Big Data, Big Consequences? Een verkenning naar privacy en big data gebruik binnen de opsporing, vervolging en rechtspraak

Lodder, A.R.; van der Meulen, N.S.; Wisman, T.H.A.; Meij, Lisette; Zwinkels, C.M.M.

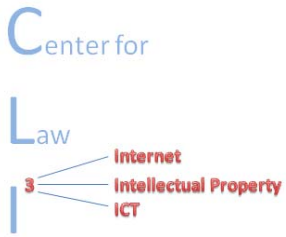
2014

document version
Publisher's PDF, also known as Version of record

BIG DATA, BIG CONSEQUENCES?

EEN VERKENNING NAAR PRIVACY EN BIG DATA GEBRUIK BINNEN DE OPSPORING, VERVOLGING EN RECHTSPRAAK

Arno R. Lodder
Nicole S. van der Meulen
Tijmen H.A. Wisman
Lisette Meij
Cees M.M. Zwinkels



Colofon

Auteurs
prof. mr. Arno R. Lodder
dr. Nicole S. van der Meulen
mr. Tijmen H.A. Wisman
Lisette Meij
mr. Cees M.M. Zwinkels CPC MPC

Uitgave
Vrije Universiteit
Faculteit Rechtsgeleerdheid, Afdeling Transnational Legal Studies
CLI^3 - Centre for Law & Internet, Intellectual Property, ICT
De Boelelaan 1105
1081 HV Amsterdam

Opdrachtgever
WODC, Ministerie van Veiligheid en Justitie
Schedeldoekshaven 131
2511 EM Den Haag

Datum
mei 2014

He was found by the Bureau of Statistics to be
One against whom there was no official complaint,
And all the reports on his conduct agree
That, in the modern sense of an old-fashioned word, he was a saint,
For in everything he did he served the Greater Community.
Except for the War till the day he retired
He worked in a factory and never got fired,
But satisfied his employers, Fudge Motors Inc.
Yet he wasn't a scab or odd in his views,
For his Union reports that he paid his dues,
(Our report on his Union shows it was sound)
And our Social Psychology workers found
That he was popular with his mates and liked a drink.
The Press are convinced that he bought a paper every day
And that his reactions to advertisements were normal in every way.
Policies taken out in his name prove that he was fully insured,
And his Health-card shows he was once in hospital but left it cured.
Both Producers Research and High-Grade Living declare
He was fully sensible to the advantages of the Instalment Plan
And had everything necessary to the Modern Man,
A phonograph, a radio, a car and a frigidaire.
Our researchers into Public Opinion are content
That he held the proper opinions for the time of year;
When there was peace, he was for peace: when there was war, he went.
He was married and added five children to the population,
Which our Eugenist says was the right number for a parent of his generation.
And our teachers report that he never interfered with their education.
Was he free? Was he happy? The question is absurd:
Had anything been wrong, we should certainly have heard.

W. H. Auden, 1938 T*he Unknown Citizen*

Number Six: Where am I?
Number Two: In the Village.
Number Six: What do you want?
Number Two: Information.
Number Six: Whose side are you on?
Number Two: That would be telling. We want information… information… information.
Number Six: You won't get it.
Number Two: By hook or by crook, we will.
Number Six: Who are you?
Number Two: The new Number Two.
Number Six: You are Number Six.
Number Six: I am not a number! I am a free man!

Opening van iedere aflevering van de Britse TV serie The Prisoner, uit 1967

Inhoudsopgave

Management samenvatting

We are building a new digital society, and the values we build or fail to build into our new digital structures will define us. Critically, if we fail to balance the human values that we care about, like privacy, confidentiality, transparency, identity and free choice with the compelling uses of Big Data, our Big Data Society risks abandoning these values for the sake of innovation and expediency.

Richards & King (2014)

Het lijkt de gouden graal van de informatiesamenleving: uit een grote berg ongestructureerde informatie allerhande niet voorziene verbanden en samenhang ontdekken. Aan de hoeveelheid informatie hoeft het niet te liggen, die is er in overvloed. De mogelijkheden van de technologie, zowel qua opslag als rekencapaciteit, vormen ook steeds minder een belemmering. Niets lijkt aan een glorieuze toekomst van Big Data analysis in de weg te staan.

Ook binnen het domein van veiligheid en justitie zijn er mogelijkheden. De taak van juristen is om de randvoorwaarden aan te geven waarbinnen de mogelijkheden van de technologie kunnen worden benut. In een democratische samenleving is het van belang dat burgers de overheid vertrouwen. Door de recente onthullingen omtrent de activiteiten van veiligheidsdiensten lijkt er van een kentering sprake.

In deze verkenning is ingegaan op de privacy aspecten van Big Data analysis binnen het domein Veiligheid en Justitie. Besproken zijn toepassingen binnen de rechtspraak zoals voorspellen van uitspraken en gebruik in rechtszaken. Met betrekking tot opsporing is onder andere ingegaan op predictive policing en internetopsporing. Na een uiteenzetting van de privacynormen en toepassingsmogelijkheden, zijn de volgende zes uitgangspunten voor Big Data toepassingen voorgesteld:

1. Bepaal te analyseren probleem en definieer doel voor verwerking
2. Selecteer data en beperk verzamelen
3. Bewaar niet langer dan noodzakelijk
4. Wees transparant
5. Beveilig informatie
6. Evalueer de uitkomsten kritisch

1 Inleiding

Customs law had a pre-digital focus which, when applied to the technical age, did not take into account the amount of personal information or the frequency of use.

Fisher (2013)

Zeker na het in 2013 verschenen boek *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (*Mayer-Schonberger & Cukier 2013)* staat 'Big Data' prominent op wetenschappelijke en beleidsagenda's.

Binnen justitie zijn er op dit moment al Big Data toepassingen, zoals de Web Voyager, die grote hoeveelheden data van het internet analyseert op zoek naar verdachte patronen. Door berichtgeving over PRISM wordt de vraag waar de grenzen van verwerkingen van Big Data liggen nadrukkelijker gesteld dan voorheen. Deze grenzen zijn uit de aard van de te verrichten activiteiten bij veiligheidsdiensten minder scherp dan bij de opsporing en vervolging.

Het gebruik van Big Data biedt steeds meer kansen om voorspellingen te doen over allerlei onderwerpen. Naast het voorspellen van menselijk gedrag op individueel niveau of het reconstrueren van individuele gebeurtenissen middels Big Data, is het ook mogelijk op meer algemeen (op een hoger aggregatieniveau) voorspellingen en/of uitspraken te doen over (sociale of criminele) fenomenen op maatschappelijk niveau. Big Data worden dan niet voor opsporingsdoeleinden gebruikt, maar voor algemene inzichten. Dergelijke inzichten over mogelijk crimineel gedrag of een grotere kans daarop, kunnen van belang zijn voor het algemene veiligheidsbeleid. Naarmate Big Data op een abstracter niveau worden geanalyseerd, zijn de mogelijke inbreuken op de persoonlijke levenssfeer minder voor de hand liggend. Toch moet niet veronachtzaamd worden dat ook in deze informatie bepaalde personen of groepen herkenbaar kunnen zijn. Ook bij abstracte

analyses moet goed nagedacht worden over wat de impact op de persoonlijke levenssfeer kan zijn van zowel de analyse als de daaruit volgende resultaten. Een van de keerzijden van aggregeren kan bijvoorbeeld stigmatisering van mensen of groepen zijn. De vraag die bij het inzetten van Big Data toepassingen moet worden gesteld is in hoeverre het recht (jurisprudentie en regelgeving) de maatschappelijke ontwikkelingen stimuleert of juist afremt en wat de gewenste rol en verhouding is.

Deze achtergrond gaf aanleiding voor het WODC van het Ministerie van Veiligheid en Justitie een verkennend onderzoek te laten uitvoeren naar de juridische randvoorwaarden, in het bijzonder inzake privacy, voor de toepassingen van Big Data binnen het domein Veiligheid en Justitie, in het bijzonder opsporing en vervolging alsmede geschillenoplossing. Het Centre for Law & Internet, Intellectuele Eigendom, ICT (CLI^3) te Amsterdam heeft in opdracht van het WODC dit onderzoek uitgevoerd, in de periode december 2013 tot april 2014.

2 Doelstelling en vraagstelling

De doelstelling van dit onderzoek is inzicht te bieden in met welke juridische uitgangspunten, met name inzake privacy, rekening moet worden gehouden bij de inzet van Big Data toepassingen binnen het domein Veiligheid en Justitie, in het bijzonder opsporing en vervolging alsmede geschillenoplossing. Daarbij moet vanuit juridisch perspectief duidelijk worden gemaakt welke knelpunten en kansen zich voordoen bij Big Data toepassingen. De vraagstelling die centraal staat is:

> Welke juridische en met name privacyrechtelijke uitgangspunten dienen in acht te worden genomen bij de inzet Big Data toepassingen binnen het domein Veiligheid en Justitie teneinde de mogelijkheden die deze technologie biedt optimaal te benutten?

2.1 Afbakening

Deze studie heeft een verkennend karakter naar een nieuw fenomeen dat nog niet systematisch in kaart is gebracht. Daarom is gekozen voor een korte schets van mogelijk relevante aspecten. De privacyrechtelijke aspecten worden nader uitgediept voor zover deze direct raken aan de inzet van Big Data toepassingen. De gevallen waarin privacy vragen opwerpt die niet specifiek te relateren zijn aan Big Data toepassingen wordt niet nader op ingegaan, dan wel worden deze aspecten kort aangestipt indien dit noodzakelijk is voor een coherente en consistente behandeling van de problematiek. Bij de behandeling van inzet van Big Data toepassingen binnen het domein Veiligheid en Justitie beperken we ons tot geschillenoplossing, met name rechtspraak en de opsporing.

2.2 Methoden van onderzoek

Vanwege het verkennende karakter is aanvankelijk gekozen voor een combinatie van literatuuronderzoek en interviews met deskundigen. Het literatuuronderzoek is gebaseerd op wetenschappelijke publicaties, beleidsrapporten van overheden, organisaties en bedrijven, alsmede te valideren of anderszins betrouwbare media- en internetberichten.

Wat betreft de interviews is in aanvang een representatieve vertegenwoordiger uit de kringen van de opsporing alsmede rechtspraak ondervraagd. Binnen de rechtspraak bleek er in het geheel nog niet nagedacht te zijn over mogelijke inzet van Big Data toepassingen, althans niet volgens degene die op aanraden van binnen justitie en de Raad voor de Rechtspraak geraadpleegde specialisten het meest in aanmerking kwam. Dit is overigens niet onbegrijpelijk, aangezien de rechtspraak op dit moment nog in een transitie fase verkeerd richting elektronische rechtspraak. Binnen de opsporing bevestigde het interview het beeld dat bestond. Daarnaast zijn een groot aantal technische deskundigen bij gelegenheid kort bevraagd over het fenomeen Big Data toepassingen. Wat daarbij opviel was de veelal gehoorde opmerking dat nog niet geheel duidelijk is wat Big Data toepassingen behelzen, anders dan het inzetten van algoritmen om ongestructureerde dataverzamelingen die te groot zijn om te analyseren met klassieke technieken mogelijk is. Aangezien er een aanpalend onderzoek in opdracht van het WODC is uitgevoerd hebben we de technische invalshoek beperkt tot het op een voor een ieder begrijpelijke manier omschrijven wat Big Data toepassingen zijn en welke mogelijkheden deze bieden.

2.3 Opbouw

Het rapport is opgebouwd als volgt. In hoofdstuk 2 gaan we kort in op de vraag wat Big Data en Big Data analysis is. Hoofdstuk 3 schetst in hoofdlijnen de regulering rond privacybescherming. In hoofdstuk 4

worden de praktische en juridische merites van enkele mogelijke toepassingen van Big Data analysis rond de rechtspraak behandeld, gevolgd door een uiteenzetting in hoofdstuk 5 rond Big Data analysis binnen de opsporing. Voortbouwend op hetgeen in hoofdstuk 2-5 behandeld is, worden in hoofdstuk 6 uitgangspunten voor de toepassing van Big Data analysis voorgesteld.

3 Big Data en Big Data analysis

An acceptable level of information flow into Big Data analysis is one that yields acceptable tradeoffs between risks and benefits. The problem is to find a level of information flow that does that.

Sloan & Warner (2013)

De verwachting is dat we in 2020 zo'n 35 zettabyte aan data hebben opgeslagen. Dit staat gelijk aan een stapel dvd's met data, die opgestapeld tot halverwege de planeet Mars reikt. Om overweg te kunnen met zeer grote, ongestructureerde, niet relationele dataverzamelingen is meer en andere kennis nodig dan tot nu toe gebruikt werd. Denk aan standaarden, filters, analysetechnieken, metadata, opslagtechnieken, zoektechnieken, beveiliging, en het beschermen van gegevens en sector- of branche specifieke databewerkingen. Bovendien is er veel (meer dan in het verleden) rekenkracht nodig om bewerkingen uit te voeren, niet in de laatste plaats omdat analyses steeds vaker real-time worden gedaan (real-time analytics). De technologie die in dit licht het meest belovend is, is Big Data analysis.

Als we de technologie-blogs, Google ads en marketing- en ICT-consultants mogen geloven is Big Data "the next big thing". Hellerstein (2008) kondigde deze ontwikkeling aan als *The Industrial Revolution of Data*. Data is volgens hem de motor van ongekende bedrijfseconomische en maatschappelijke mogelijkheden. Vier jaar later is Thiele (2012) bijzonder stellig over de mogelijkheden van Big Data en dat we slechts aan het begin van de ongekende mogelijkheden staan:

> "Big Data today, is what the web was in 1993. We knew the web was something and that it might get big, but few of us really understood what "big" meant. Today we aren't even scratching the surface of the Big Data opportunity".

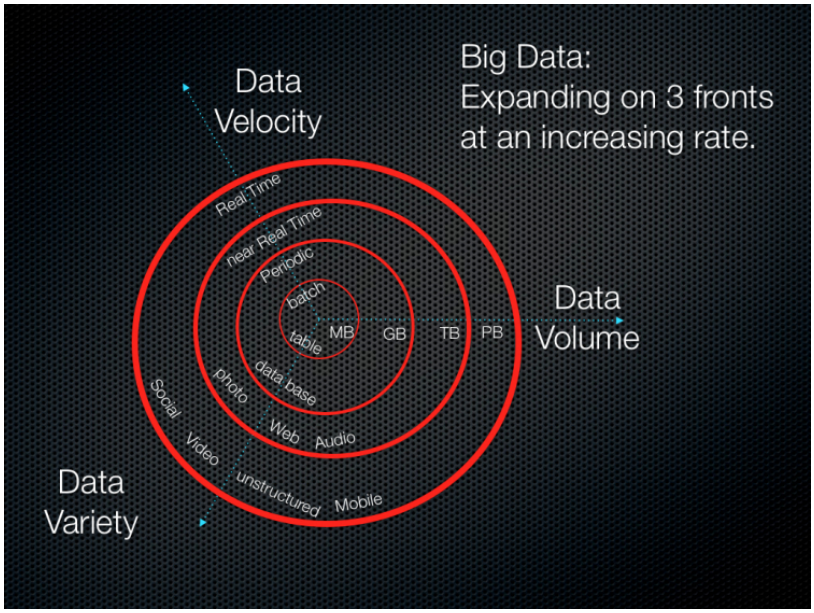
We zullen in dit hoofdstuk ingaan op wat onder Big Data verstaan moet worden en wat Big Data analysis inhoudt.

3.1 Big Data

Big Data brengt in de kern niet meer of minder tot uitdrukking dan dat er heel veel data zijn, een niet te bevatten hoeveelheid gegevens op het internet alsmede daarbuiten die iedere seconde uitbreidt. Niet alleen is veel informatie op internet te vinden, ook de internet en telecommunicatie verkeersgegevens alsmede sensoren op gebouwen, bruggen, wegen, etc. genereren enorme hoeveelheden data. Daarnaast beschikt de overheid over informatie in een groot aantal databases (kernregistraties en basisregistraties) die kunnen worden gekoppeld aan en gecombineerd met de hierboven genoemde gegevens. Tenslotte zal naar het zich laat aanzien door de al ingezette ontwikkeling van het internet naar het internet van dingen nog meer informatie over tal van objecten beschikbaar komen. Hierbij kan gedacht worden aan informatie over het huis (via o.a. slimme meters), de auto (o.a. door het per 2015 verplichte e-call systeem) en dagelijkse gebruiksvoorwerpen als elektronische tandenborstels. De hoeveelheid beschikbare data wordt vrijwel onuitputtelijk.

Zoals wel vaker met (technische) begrippen ontbreekt een algemene definitie van wat Big Data precies omvat, maar er is wel consensus over drie relevante kenmerken:

1. Volume;
2. Variety;
3. Velocity.



BRON: http://www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data

De term volume spreekt voor zich, het gaat bij Big Data om heel veel en steeds meer data. Dan is er een grote verscheidenheid in het soort gegevens: tekst, plaatjes, sensor data, etc. Tenslotte worden de data met grote snelheid verwerkt. Aanvankelijk in batch en inmiddels steeds vaker (bijna) real time. Deze laatste eigenschap ziet dus deels ook op de analyse, maar gaat daar uiteraard ook aan vooraf. Zeer snelle, real time analyse heeft alleen dan zin als er ook met grote snelheid data beschikbaar komen. Het is de combinatie van deze drie eigenschappen die bepalend zijn voor de technologie die gebruikt wordt voor de analyse. Er worden ook wel andere eigenschappen genoemd, maar de bovengenoemde drie V's geven in de kern aan wat onder Big Data verstaan moet worden. Volgens Grimes (2013) moet dan ook niet ingegaan worden op mogelijk andere aspecten van Big Data:

> ""The three V's -- volume, velocity and variety -- do a fine job of defining Big Data. Don't be misled by the "wanna-V's:" variability, veracity, validity and value."

Toch wordt steeds vaker wordt ook een van de "wanna-V's" als vierde V aan het rijtje toegevoegd, namelijk: Veracity. Hierbij gaat het erom of de data nuttig is voor een specifieke analyse. De vraag is dan of de gegevens die worden ingevoerd kunnen leiden tot zinvolle uitkomsten. Deze vierde V wordt ook wel gevat onder de termen Value en Relevance. In sommige gevallen wordt de verzameling V's zelfs uitgebreid met (5) Validity en (6) Volatility (Normandeau 2013). Hierbij gaat het erom of de data relevant en correct zijn. Met Grimes zijn wij van mening dat de aanvullende V's niet echt iets toevoegen. Ook de term *Complexity* wordt wel gebruikt om het fenomeen Big Data mee te duiden. Complexiteit is wederom niet echt kenmerkend voor Big Data. Zo is het werken met grote, gecompliceerde databestanden op zichzelf niet nieuw en ook binnen het WODC heeft men de mogelijkheid om grote hoeveelheden data te bewerken, te ordenen en te onderzoeken op samenhang. Een duidelijk voorbeeld van de hoeveelheid data, de verscheidenheid daarvan en de snelheid waarmee ze ontstaan vormt het project Global Pulse. De Verenigde Naties is in 2009 begonnen met dit project bedoeld om met behulp van Big Data voorspellingen te doen over de gevolgen van economische crises waardoor tijdig en juist kan worden ingegrepen:[1]

> "in response to the need for more timely information to track and monitor the impacts of global and local socio-economic crises"

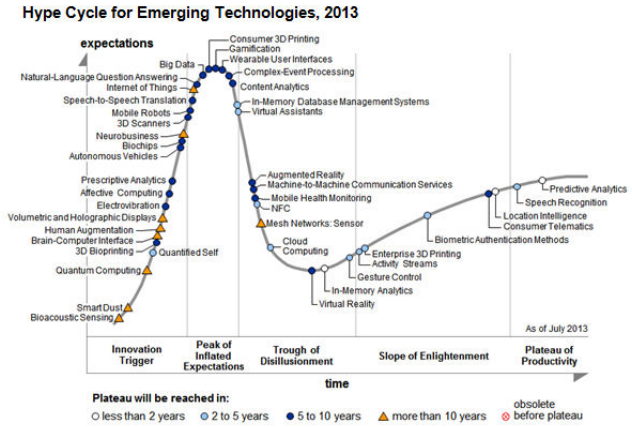
Global Pulse maakt gebruik van Big Data via:[2]

[1] http://www.unglobalpulse.org/
[2] http://www.unglobalpulse.org/about-new

1.Online Sources - Public news stories, blogs, Twitter, Facebook, obituaries, birth announcements, job postings, e-commerce, etc.
2.Private Sector Partnerships - Anonymized data from telecommunications companies, mobile banking, online searches, hotline usage, transit companies etc.
3.Physical Sensors - Satellite imagery, video, traffic sensors, etc.
4.Crowdsourced Reports - Information actively produced or submitted by citizens through mobile phone-based surveys, user generated maps, etc.

3.2 Big Data Analysis

Op onderstaande, regelmatig in presentaties gebruikte, grafiek van Gartner over de ontwikkeling van technische hypes is Big Data momenteel bovenin de hype cycle te vinden.[3]



Source: Gartner August 2013

[3] Deze cyclus loopt tot op zekere hoogte parallel met stage theory, vgl. Piaget (1970) en Nolan (1973).

Deze grafiek beoogt het verloop van hypes rond technologie weer te geven. In het begin van de grafiek is van technische innovatie sprake, maar zijn de precieze mogelijkheden nog niet helder. Dan beginnen succesverhalen de ronde te doen en komt er veel aandacht voor het verschijnsel dat echter nog niet breed wordt opgepikt. In de volgende fase neemt de teleurstelling de overhand door enkele mislukkingen en gebrek aan goede ideeën hoe de technologie te gebruiken. Daarna begint de langzame opleving, stap voor stap, doordat er meer inzicht komt in de technologie en succesvolle toepassingen zich prominenter aandienen. Tenslotte bereikt de technologie de grote massa en start ook winstgevende toepassingen, mede door grote bedrijven.[4]

Een hoog hype gehalte betekent dus geenszins dat een betreffend fenomeen zonder betekenis is, maar enkel dat er veel aandacht voor is op een bepaald moment, mogelijk meer dan op basis van nuchtere verwachtingen gerechtvaardigd is. Hoewel dit niet voor alle hypes geldt, wordt de technologie veelal in de loop der tijd dus een geaccepteerd en gewaardeerd onderdeel van de samenleving. De verwachting is dat Big Data niet snel naar de achtergrond zal verdwijnen. In een informatiesamenleving is de belangrijkste grondstof immers informatie en Big Data raakt daarmee aan de kern van onze tegenwoordige maatschappij. Of zoals door sommige wordt aangegeven er een nieuwe fase in onze samenleving is aangebroken waarin Big Data centraal staan, vergelijkbaar met de industriële revolutie (Mehta 2011) en de overgang naar de informatiemaatschappij, is op dit moment niet met zekerheid te stellen.

Grote hoeveelheden gegevens zijn op zichzelf niet interessant, maar kunnen dat worden door de juiste analyses uit te voeren. Of zoals Viergever & Koëter (2012) aangeven:

[4] Vgl. ook beschrijving van Zittrain (2008) over het internet.

> "Informatie die voorheen diep was weggestopt in bergen ongestructureerde data waar niemand naar omkeek, wordt nu toegankelijk."

Hiermee komt tot uitdrukking dat Big Data[5] op zichzelf nog weinig betekenis heeft, maar dat door analyse er informatie uit de Big Data afgeleid kan worden. Het afleiden van nieuwe, betekenisvolle informatie uit bestaande informatie is niet nieuw, en de behoefte aan goede technologie om deze analyses uit te voeren evenmin. In de jaren negentig begon de toename van data al dermate grote vormen aan te nemen dat handmatige analyses[6] niet langer mogelijk waren (Fayyad *et al*. 1996):

> "There is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information (knowledge) from the rapidly growing volumes of digital data. These theories and tools are the subject of the emerging field of knowledge discovery in databases (KDD)."

Fayyad *et al*. (1996) delen knowledge discovery op in vijf fasen:

1. data selection,
2. data pre-processing,
3. data transformation,
4. data mining and
5. interpretation.

[5] De term Big Data wordt veelal *ook* gebruikt om mogelijke toepassingen mee aan te duiden.

[6] Stranieri & Zeleznikow (2006): "Data is now collected in a variety of commercial and scientific fields in such quantities that the problem of automating the elicitation of meaningful knowledge from data has become pressing. For example, data sets from astronomical observations were once manually scanned by experts searching for anomalies or interesting patterns (…) manual analysis of data in astronomy is no longer feasible since data sets in this field often exceed many thousands of millions of records.

Er is een belangrijk verschil tussen de Knowledge Discovery in Databases en Big Data analyse. Informatie in een database is gestructureerd opgeslagen door gebruik van velden, rijen, kolommen. Voor Big Data is karakteristiek dat informatie ongestructureerd is en in allerlei verschillende bestandstypen kan zijn opgeslagen, zoals tekstbestanden, spreadsheets, powerpoint-presentaties, belgegevens, camerabeelden, etc. Begin deze eeuw verzamelde Google zoveel data dat deze niet met klassieke databases en analyse tools geanalyseerd konden worden. Eén van de eerste en bekendste ontwikkelaars van Big Data analysis software is Hadoop,[7] maar ook Oracle, die aanvankelijk geen belang stelde in de door Google gestelde vraag naar nieuwe technieken, is inmiddels actief op de Big Data markt.[8] Een ander voorbeeld is Alteryx dat zich onder andere toelegt op het op de juiste wijze vermengen van allerhande data.[9] Voor deze verkenning voert een uiteenzetting van de verschillende aanbieders en de door hen gebezigde technieken te ver en verwijzen wij onder andere naar het al eerder genoemde WODC onderzoek waar specifiek op Big Data technieken wordt ingegaan.

De stappen die bij Big Data analysis genomen worden komen grotendeels overeen met de hierboven genoemde fasen bij KDD en verloopt in grote lijnen als volgt.[10] Om te beginnen wordt bepaald welk probleem onderzocht gaat worden. Vervolgens wordt nagedacht over welke data daarvoor gebruikt kunnen worden. Op deze data worden dan de analytische tools toegepast. Als het analyse

[7] http://hadoop.apache.org/
[8] http://www.oracle.com/us/technologies/big-data/index.html
[9] https://gallery.alteryx.com/#!
[10] Zie meer uitgebreid bijvoorbeeld http://www.ikanow.com/8-proven-steps-to-starting-a-big-data-analytics-project/ 1. Problem. 2. Impact. Understand how these problems impact your business and then develop use case(s). 3. Success criteria. How will you measure the success? 4. Value & Impact. What you need to clearly understand is if this problem was solved, what would it mean for your organization? If you can't clearly define and articulate steps 1-4, there is no point in moving to step 5. Cloud or On-Premise. 6 Data requirements. 7 Identify gaps.

proces succesvol verlopen is, dan behelzen de in de gegevensset gevonden verbanden een oplossing voor het probleem. De beschreven procedure is iteratief, dus kan totdat uitkomsten gevonden zijn die voor de gebruiker nuttig zijn herhaald en verbeterd worden op ieder van de genoemde punten (probleemanalyse, dataselectie, gebruikte analyse tools).[11]

3.2.1 Succesvolle toepassingen

In het bedrijfsleven is op het meest abstracte niveau de vraag hoe meer winst gemaakt kan worden, die in allerlei kleinere deelvragen kan worden opgesplitst als hoe bereik ik mijn klanten, hoe krijg ik meer klanten, hoe kan ik mijn producten en diensten verbeteren, etc. Hiervoor kunnen verschillende binnen het bedrijf al aanwezige gegevens gecombineerd worden met bijvoorbeeld op het internet aanwezige informatie. Toepassing van analytische software moet dan vervolgens leiden tot oplossingen. Als de overheid gebruik maakt van Big Data toepassingen zal de kwaliteit van de dienstverlening, maar zeker ook de efficiëntie en daarmee gepaard gaande eventuele kostenvermindering centraal staan.

Het gebruik van Big Data is veelbelovend, maar de praktijk (het implementeren van Big Data toepassingen) is vooralsnog weerbarstig. Coleman Parks heeft in opdracht van Iron Mountain in 2012 onderzoek uitgevoerd onder 760 informatiemanagers in Europa. Dit onderzoek laat zien dat veel bedrijven worstelen met de enorme hoeveelheid data die te groot is om op een efficiënte manier te verwerken of op een betekenisvolle manier in te zetten. Hoewel bedrijven het dus lastig vinden zakelijk voordeel te behalen uit Big Data, zijn er ook succesverhalen. Zo is het autobedrijf Hertz door Big Data analysis erin geslaagd om “better focus on improvements that

[11] Zie ook http://www.sas.com/en_us/news/sascom/2012q4/big-data-delivery.html “Seven steps necessary for realizing the full potential of Big Data: **Collect, Process, Manage, Measure, Consume, Store, Govern**

our customers care about".[12] Het technologie bedrijf Capgemini bericht op hun site: "Investeringen in Big Data leveren meer resultaat op dan investeringen in het verbeteren van backoffice processen" en "Gemiddeld verbetert Big Data de businessperformance bij respondenten met 26% met zicht op verdere stijging." Een heel bekend Big Data succesverhaal is de toepassing door Amazon:[13]

> "Amazon uses Big Data also to offer a superb service to its customers. (…) They can do this because they use all the data they have collected from their customers to build and constantly improve the relationship with its customers. This is something many e-tailers can learn from."

Er zijn ook minder succesvolle experimenten. Zo plaatste een bedrijf in London vuilnisbakken die WIFI-signalen van passerende mensen opving. Deze signalen werden gebruikt om passende reclame weer te geven op de prullenbak. Zo konden ze de locatie van een passant opslaan en constateren als hij bijvoorbeeld 20 minuten in de McDonalds verbleef, om vervolgens Burger King erop te attenderen reclame te maken. In zekere zin mosterd na de maaltijd, maar wellicht dat dergelijke reclame-uitingen van invloed zijn op toekomstige hamburger-consumptie. Toen bekend werd dat de persoonlijke informatie van passanten werd gebruikt, besloot de plaatselijke overheid de prullenbakken wegens privacy inbreuken direct te laten verwijderen.

Binnen de overheid zijn er ook al Big Data succesverhalen. Zo heeft Obama zijn herverkiezing mede te danken aan het gebruik van Big Data. Door Big Data analyse kon hij precies uitzoeken waar de

---

[12] http://www-01.ibm.com/software/ebusiness/jstart/portfolio/hertzCaseStudy.pdf "Hertz gathers an amazing amount of customer insight daily, including thousands of comments from web surveys, emails and text messages. We wanted to leverage this insight at both the strategic level and the local level to drive operational improvements," said Joe Eckroth, Chief Information Officer, the Hertz Corporation.

[13] http://www.bigdata-startups.com/BigData-startup/amazon-leveraging-big-data/

voorkeuren van de kiezers lagen en hier op inspelen (Magyar 2013). In vervolg hierop werd in maart 2012 aangekondigd dat 6 federale departementen voor niet minder dan $200 miljoen in Big Data investeren.[14] Meer dan een jaar later roepen ze alle mogelijke personen die een bijdrage kunnen leveren zoals aandeelhouders op om mee te werken aan hun Big Data initiatief.[15] De National Science Foundation (NSF) roept personen en bedrijven op hun eigen Big Data project of initiatief in te zenden.[16] The Network and Information Technology Research and Development (NITRD) Program heeft vanuit de overheid de Big Data Senior Steering Group opgestart en organiseren workshops en evenementen om zo de groei van Big Data onderzoek te stimuleren in samenwerking met andere organisaties.[17]

3.2.2 Verzamelen en Profielen

Twee aspecten die vanuit het perspectief van privacy nadrukkelijk aandacht vragen en we daarom later op terugkomen, zijn het verzamelen van gegevens en het na analyses opstellen van profielen. Bij Big Data analyse moet nagedacht worden over welke data gebruikt kunnen worden om het probleem te adresseren, waarbij in beginsel geldt dat hoe meer gegevens verzameld worden des te beter de uitkomsten zijn. Dit betekent dat in eerste instantie data relatief onbeperkt verzameld wordt en deze pas tijdens de analyse wordt geschift. Dit kan spanning opleveren met onder andere het doelbindingsbeginsel dat bepaalt dat duidelijk omschreven moet worden voor welk doel gegevens verzameld worden, zeker in combinatie met het data limiteringbeginsel. Dit laatste beginsel bepaalt dat niet meer data moeten worden verzameld dan voor het beoogde doel noodzakelijk is.

---

14 http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf

15 http://www.whitehouse.gov/blog/2013/04/18/unleashing-power-big-data

16 http://www.nsf.gov/cise/news/2013-BIGDATA-announcment.jsp

17 http://www.nitrd.gov/nitrdgroups/index.php?title=Big_Data_(BD_SSG)#title

Het andere aspect is de profielen die na Big Data analyse worden opgesteld. Dit zijn in de regel abstracte, niet naar een specifieke persoon te herleiden verzamelingen gegevens die bepaalde eigenschappen vertegenwoordigen. Informatie over gedrag, locatie en emotie kan worden gebruikt om profielen op te stellen. In marketing kringen bestaat het idee dat het door profielen mogelijk is om klanten op de hoogte te brengen van precies dat wat ze willen, ook als ze dat zelf nog niet weten. Behalve dat deze profielen alleen opgesteld kunnen worden door informatie over personen te analyseren en te aggregeren, kunnen de opgestelde profielen vervolgens worden toegepast op individuele personen en worden er soms (ingrijpende) gevolgtrekkingen aan worden.

3.2.3 Dat (correlatie) vs. Waarom (causatie)

Een ander punt dat zeker bij toepassingen binnen de overheid aandacht vraagt is de waarde die wordt gehecht aan de verbanden die door toepassing van speciaal ontwikkelde analytics software worden gevonden. Het leidende mantra is immers DAT er verband is tussen de data,[18] de correlatie wordt blootgelegd. Aan WAAROM er een verband is, dus de causaliteit, wordt niet of veel minder belang aan gehecht.

Hierin schuilt een gevaar van Big Data toepassingen. Als voor een supermarkt niet duidelijk zou zijn WAAROM het plaatsen van een kratje bier naast luiers een grotere bieromzet oplevert, doet dat er niet veel toe. Van belang is uiteraard het gegeven DAT er meer omzet wordt gegenereerd. Als op basis van Big Data analyse echter een vermoeden van een criminele organisatie ontstaat door een verband tussen kinderporno, vuurwapens, cyberwar, Arno Lodder en Wouter Stol, dan zou dit voetstoots als juist kunnen worden aangenomen (DAT) maar is interpretatie (WAAROM) in het belang van twee laatstgenoemden. De reden voor de correlatie is immers niet dat zij al

[18] Dit komt onder andere nadrukkelijk naar voren in Mayer-Schonberger & Cukier (2013).

dan niet in gezamenlijkheid vergelijkbare criminele activiteiten verrichten, maar respectievelijk dat ze samen in 2008 een WODC-onderzoek over kinderporno uitvoerden, Lodder in 2012 een hoofdstuk over cyberwar voor een boek van Stol schreef en ze beiden in 2013 op het Europese hand vuurwapen congres een lezing verzorgden. Dit voorbeeld lijkt wellicht vergezocht, maar een waar gebeurde situatie is een jongetje dat na de liquidatie van Osama Bin Laden zich zorgen maakte over Obama en op het internet daarom de Amerikaanse president waarschuwde op te passen voor aanslagen. Binnen een uur werd er door de veiligheidsdienst aangebeld.[19] Een Nederlands voorbeeld is de Fortuna Sittard fan die met zijn vader via WhatsApp over een bom communiceerde zonder er daadwerkelijk over te beschikken dan wel dit onderwerp serieus te adresseren.[20] Ook hier waren agenten vrij snel ter plaatse om tot eventuele arrestatie over te gaan.

Uit deze gevallen komt naar voren dat de geheime dienst en politie constant een uitzonderlijke grote datastroom op het internet alsmede afkomstig van telecommunicatie monitort en filtert. Ook illustreren deze gevallen dat op basis van risicoprofielen wordt besloten tot handelen over te gaan.

3.3 Werkdefinitie

[19] Te zien in de documentaire *Terms And Conditions May Apply* (2013) van Cullen Hoback.

[20] Na Kamervragen ontkende de politie overigens WhatsApp verkeer te monitoren (NRC 10 oktober 2013): “Volgens Chris Timmermans, woordvoerder van de politie Limburg, was het politiebezoek een reactie op een mondelinge tip, die niets te maken heeft met het WhatsApp-bericht. De politie tapt WhatsApp-verkeer niet zomaar af, zegt hij.” Uit betrouwbare bron vernamen wij echter over door de politie gebruikte software: “het is aangeprezen en gekocht voor monitoren van verdachten, maar langzamerhand is iedereen na gebruik van een woord uit de verdachte woordenlijst verdacht tenzij tegendeel achteraf is bewezen.”

Binnen de Europese Unie leeft het idee dat er meer gebruik van Big Data gemaakt zou moeten worden. Zo constateerde Neelie Kroes in november 2013 dat van de 20 grootste Big Data bedrijven er maar 3 uit de Europese Unie komen. Informatie uit de gezondsheidszorg wordt veelal gezien als een nuttige toepassing van Big Data analysis. Kroes ziet ook onder andere mogelijkheden in de gezondheidszorg, net als de Amerikanen:

> "The healthcare industry could see big benefits from Big Data – including improvements in drug trial safety, disease surveillance, prescribed treatments, and patient outcomes. Two-thirds of federal executives working in healthcare-focused agencies believe that Big Data will improve population health management and preventive care."

De stimulering voor de economie die Kroes ziet wordt in haar ogen niet gehinderd door privacy (Schoemaker 2013):

> "De meeste data in Big Data is geen persoonlijke informatie en daarom zouden we wat minder huiverig moeten zijn in het gebruik van die data."

Hoewel betoogd kan worden dat het analyseren van dergelijke informatie niet persoonsgegevens hoeft te bevatten, is gezien het gevoelige karakter van deze informatie, zeker in de gezondheidszorg, noodzakelijk dat goed nagedacht wordt over de uitvoering alsmede over welke informatie naar buiten gebracht wordt. Het is de vraag in hoeverre de informatie die ontstaat door Big Data analysis inderdaad niet persoonlijk is of kan worden. In dit rapport zullen we ingaan op de juridische consequenties van Big Data analysis in het algemeen en binnen justitie in het bijzonder en de spanning die er bestaat tussen Big Data toepassingen en met name privacy.

Er wordt wel gezegd dat bij Big Data analysis de hoeveelheid gegevens zó groot en divers is dat ze niet meer te beheren zijn met

tot nu toe gebruikelijke middelen, zoals conventionele databases. Voor de verdere bespreking gaan we van het volgende uit. Big data analysis gaat om het verwerken en analyseren van grote hoeveelheden gegevens: data volumes. Deze grote hoeveelheid gegevens zijn niet uniform of gestructureerd, maar gevarieerd, ongeordend, en in verschillende dataformaten. Tenslotte vindt de verwerking met grote snelheid, in sommige gevallen zelfs “on the fly” plaats.

4 Normen rondom Privacybescherming

(…) system of "notice and consent". In the era of Big Data, however, when much of data's value is in secondary uses that may have been unimagined when the data was collected, such a mechanism to ensure privacy is no longer suitable.

Mayer-Schonberger & Cukier (2013), p. 173

De WP29 verwoordt kernachtig de mogelijkheden en risico's:[21]

> "With all its potential for innovation, big data may also pose significant risks for the protection of personal data and the right to privacy."

Europese regels omtrent privacy worden momenteel herzien. De Nederlandse Wet bescherming persoonsgegevens is sterk verouderd. De aan deze wet ten grondslag liggende EU Privacyrichtlijn uit 1995 is ontworpen in de tijd dat internet niet of nauwelijks een rol van betekenis speelde in de samenleving. Er zijn in deze richtlijn echter beginselen vervat die voor privacy onverminderd relevant zijn en die ook uitdrukking geven aan de waarborgen waarmee het recht op privacy is omkleed in het Europees Verdrag voor de Rechten van de Mens (EVRM). Het is op dit moment niet zeker of de begin 2012 voorgestelde algemene verordening gegevensbescherming (AVG) nog voor het einde van de termijn (2014) van de huidige Commissie en Europees Parlement wordt vastgesteld. Op 12 maart 2014 is de Verordening door het Europese parlement aangenomen. Het Parlement is er van overtuigd dat de verordening zal worden vastgesteld,[22] maar vanwege de co-decisie procedure moet de Raad

[21] WP29, Opinion 03/2013 on purpose limitation, 2 april 2013, p. 35.

[22] Progress on EU data protection reform now irreversible following European Parliament vote: "Today's plenary vote means the position of the Parliament is now set in stone and will not change even if the composition of the Parliament changes following the European elections in May." http://europa.eu/rapid/press-release_MEMO-14-186_en.htm

van Ministers nog akkoord gaan. Deze hebben zich eerder overigens positief over de Verordening uitgelaten.

Dat het grondrecht privacy niet zo snel zal verdwijnen als door sommige gedacht[23] volgt in zekere zin uit de titels van twee boeken:

The end of Privacy
The end of Privacy, How total surveillance is becoming a reality.

Deze boeken zouden net uitgebracht kunnen zijn, maar verschenen beide 15 jaar geleden en werden geschreven door respectievelijk C.J. Sykes en R. Whitaker. Het geeft wel aan dat privacy al tenminste 15 jaar onder druk staat. In een eveneens in 1999 verschenen Nederlandse Rathenau studie werd door Smink e.a. aangegeven dat:

> "Autonomie, zelfbeschikking, integriteit, zelfstandigheid, bewegingsvrijheid, gelijkheid en vrijwaring van stigmatisering worden gezien als onderliggende waarden die privacy van een normatieve grondslag voorzien."[24]

Vooral de stigmatisering is een aspect dat bij Big Data analysis aandacht verdient. We zullen er hieronder, mede aan de hand van een voorstel in de Privacy verordening, aandacht aan besteden. In de media en het wetenschappelijk discours krijgt vooral de Privacy verordening aandacht. In het kader van dit rapport moeten we zeker ook noemen de tegelijkertijd voorgestelde Richtlijn inzake verwerking van persoonsgegevens door politie en justitie.[25]

---

[23] Zoals de vaak aangehaalde frase "Privacy is dead, get over it".

[24] Dit citaat is afkomstig uit de boekbespreking "Het einde van privacy zoals we haar kennen?" van Lynsey Dubbeld dat verscheen in *Krisis* 2001, p. 63-70 http://www.krisis.eu/content/2001-2/2001-2-06-dubbeld.pdf waarin vijf boeken worden besproken waaronder de laatste drie in de hoofdtekst genoemde.

[25] Voorstel voor een RICHTLIJN VAN HET EUROPEES PARLEMENT EN DE RAAD betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens door bevoegde autoriteiten met het oog op de voorkoming,

Dit rapport valt tussen twee privacy regimes in, waarbij de nieuwe regelgeving nog niet uitgekristalliseerd is. Behalve deze tussenfase in privacyregulering, ontbreekt in dit rapport de ruimte om uitgebreid in te gaan op bestaande of toekomstige regelgeving. Bovendien volstaat voor een goed begrip de bespreking van de juridische beginselen die aan de verwerking van persoonsgegevens ten grondslag liggen. Hierbij is het van belang kritisch naar deze beginselen te kijken.

4.1 Kern beginselen gegevensbescherming

Artikel 2 lid 2 Wbp bepaalt dat deze wet niet van toepassing is op de verwerking van persoonsgegevens ten behoeve van de uitvoering van de politietaak, noch ten behoeve van de inlichtingen- en veiligheidsdiensten. Wanneer bij Big Data analysis persoonsgegevens worden verwerkt, komt deze verwerking te vallen onder de Wet bescherming persoonsgegevens (Wbp), de Wet politiegegevens (Wpol ook wel Wpg)[26], de Wet op de inlichtingen- en veiligheidsdiensten (Wiv) of de Wet justitiële en strafvordelijke gegevens (Wjsg), afhankelijk van de instantie die de analyse uitvoert. Er bestaan overeenkomsten tussen de eisen die de Wbp en de Wpol stellen aan de verwerking van persoonsgegevens. De Wiv is wat dit betreft een vreemde eend in de bijt, omdat het praktisch geen materiële beperkingen stelt aan de gegevens die mogen worden verwerkt. In deze paragraaf worden de belangrijkste beginselen van gegevensverwerking behandeld, die zowel in de Wbp als de Wpol zijn te vinden en wordt er aangegeven hoe deze zich verhouden tot Big Data analysis.

4.1.1 Doel en grondslag

het onderzoek, de opsporing en de vervolging van strafbare feiten of de tenuitvoerlegging van straffen, en betreffende het vrije verkeer van die gegevens

[26] Zie beide officiële afkortingen via http://wetten.overheid.nl/BWBR0022463/

Big Data analysis laat de technologie verbanden tussen gegevens ontdekken. Bij Big Data analysis kan weliswaar aangegeven worden dat gegevens met dat doel gebruikt gaan worden, maar de vraag is in hoeverre een dergelijk doel voldoende concreet is. Het formuleren van een specifiek doel is bij Big Data analyse lastig, te meer daar een vereiste bij doelspecificatie is dat het doel welbepaald en uitdrukkelijk omschreven moet zijn. Het is van belang bij Big Data analysis goed na te gaan met welk doel deze technologie wordt toegepast.

In geval van Big Data analysis is het lastig om toestemming te vragen als de verantwoordelijke zelf nog niet duidelijk kan omschrijven met welk doel hij de gegevens gaat verwerken. Toch zal bij het vaststellen van het probleem waarvoor Big Data gebruikt gaan worden ook het doel van de daarvoor benodigde gegevensverwerking zo goed mogelijk moeten worden gespecificeerd.

Toestemming die personen moeten geven voor de verwerking van persoonsgegevens speelt bij de overheid minder dan bij bedrijven. Bij overheden is de grondslag om persoonsgegevens te verwerken veelal een wettelijke plicht.[27] Daarnaast zal zeker bij de uitvoering van taken van veiligheids- of opsporingsinstanties het vrijwel altijd de bedoeling zijn dat de betrokkene juist niet weet dat deze onderwerp van een onderzoek is.

Het kan voorkomen dat voor de oorspronkelijke verwerking een geldige grondslag is, maar dat de Big Data analysis een nieuwe verwerking behelst waarbij er in beginsel ook een nieuwe grondslag vereist is. Zo hebben de gegevens uit de *basisregistratie personen* tot doel overheidsorganen toegang te geven tot de in deze registratie opgenomen persoonsgegevens, voor zover deze gegevens noodzakelijk zijn voor de vervulling van hun taak (artikel 3.1 Wet

[27] Artikel 8 sub c Wbp: “de gegevensverwerking is noodzakelijk om een wettelijke verplichting na te komen waaraan de verantwoordelijke onderworpen is;”

basisregistratie personen). Deze registratie is bij de belastingdienst gekoppeld aan allerhande inkomensgegevens. De belastingdienst kan ook camerabeelden op parkeerplaatsen bij bouwmarkten en in de buurt van pretparken gebruiken om vast te stellen of iemand zijn auto wel voor de zaak gebruikt. Deze op zichzelf geldige grondslag is niet geschikt om dezelfde beelden te gebruiken om de correctheid van andere belastinginformatie te checken. Hiervoor is dan een nieuwe grondslag vereist.

Als de verwerking plaatsvindt onder het regime van de Wbp dan is er geen nieuwe grondslag nodig indien de verwerking verenigbaar is met de doeleinden waarvoor de persoonsgegevens zijn verkregen. Hierbij moet een lijst van strikte criteria in acht worden genomen, waarbij er ook moet worden gekeken naar de gevolgen van de verwerking voor de betrokkene. Wanneer deze zeer ingrijpend zijn, is het niet waarschijnlijk dat deze verwerking toelaatbaar wordt geacht. Indien de verwerking plaatsvindt onder het regime van de Wpol is het van belang dat de Big Data analysis noodzakelijk is voor de doeleinden geformuleerd in de Wpol. Daarnaast mogen politiegegevens voor een ander doel worden verwerkt wanneer de wet daar uitdrukkelijk in voorziet.

Het is van belang te realiseren dat de wetgeving niet over registers gaat, zoals bij de voorganger de Wet persoonsregistraties het geval was. De Wbp gaat over het verwerken van persoonsgegevens en legt degene onder wiens verantwoordelijkheid dit geschied bepaalde plichten op. Het verwerken is een veel ruimer begrip dan gegevens die opgeslagen liggen in een database. Het combineren van gegevens die al dan niet direct herleidbaar zijn tot een persoon vallen onder deze regelgeving.

4.1.2 Doelbinding

Het beginsel van doelbinding houdt in dat gegevens slechts mogen worden verwerkt voor welbepaalde, uitdrukkelijk omschreven en gerechtvaardigde doeleinden. Dit beginsel wordt in de Wbp anders

uitgewerkt dan in de Wpol, maar in beide wetten is het een eis voor een rechtmatige verwerking. Rechtmatig is immers alleen die verwerking waarvoor nadrukkelijk een doel omschreven is, dat bovendien legitiem moet zijn. De spanning die dit beginsel met Big Data analysis oproept is gelegen in het feit dat hierbij de doeleinden vaker niet van te voren welbepaald en uitdrukkelijk omschreven zijn. Verwerkingen die vallen onder de Wbp kunnen derhalve botsen met dit beginsel. De doeleinden genoemd binnen de Wpol zijn het voorkomen en opsporen van strafbare feiten, het handhaven van de openbare orde, het verlenen van hulp aan hen die deze behoeven en het uitoefenen van toezicht op het naleven van regelgeving. Binnen deze brede categorieën kan een analyse makkelijk worden ingedeeld en het valt daarom te verwachten dat doelbinding in het kader van opsporing niet snel tot problemen zal leiden.

4.1.3 Data minimalisatie

Een privacybeginsel dat inherent tegenstrijdig lijkt met Big Data analysis is data minimalisatie. Dit beginsel geeft uitdrukking aan de eis dat gegevens slechts worden verwerkt voor zover zij toereikend, ter zake dienend en niet bovenmatig zijn. Er mag niet meer data worden verwerkt dan noodzakelijk is voor het te realiseren doel. Deze eis geldt zowel binnen de Wbp als de Wpol. Hoe welbepaalder en uitdrukkelijker dit doel is omschreven, des te lastiger is het om aan dit vereiste te voldoen. Bovendien stelt de Wbp als eis dat gegevens niet langer mogen worden "bewaard in een vorm die het mogelijk maakt de betrokkenen te identificeren, dan noodzakelijk is voor de verwerkelijking van de doeleinden waarvoor zij worden verzameld of vervolgens worden verwerkt."[28] Dit beginsel zou er in theorie toe moeten leiden dat persoonsgegevens die worden verwerkt onder de Wbp worden verwijderd zodra ze niet meer relevant zijn voor het oorspronkelijke doeleinde waarvoor zij werden verwerkt. De praktijk laat echter een heel ander beeld zien, waarin de regel lijkt te zijn dat persoonsgegevens langer worden opgeslagen

[28] Artt. 10 lid 1 Wbp.

dan strikt noodzakelijk. Dit biedt enerzijds praktisch gezien kansen voor Big Data analysis, anderzijds is dit juridisch onrechtmatig. De overheid als hoeder van de wet moet daarom prudentie betrachten bij het gebruik van gegevens die onder de Wbp door haar worden verwerkt en die zij overweegt te gebruiken in het kader van Big Data analysis. Dit is eveneens belangrijk bij verwerkingen die vallen onder de Wpol. Hoewel de eisen die worden gesteld binnen deze wet minder strikt zijn geformuleerd, zijn het juist verwerkingen in deze sfeer die zeer gevoelig kunnen liggen en waarbij de overheid uiterste zorgvuldigheid dient na te streven. Op het belang hiervan wordt teruggekomen in de paragraaf over het EVRM.

4.1.4 Verzamelen van persoonsgegevens

Als informatie is te herleiden tot een bepaalde persoon dan is er sprake van persoonsgegevens, of zoals in de voorstel Privacy verordening verwoordt:

> any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, unique identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social or gender identity of that person;

Door de grote hoeveelheid beschikbare data kan vrij snel aangenomen worden dat er een “identifier” is, een onderdeel van de informatie die deze kan terugvoeren tot een bepaalde persoon. De bewering dat een verzameling gegevens geen persoonsgegevens bevatten is daarom steeds lastiger staande te houden.

Voor meerdere interpretaties vatbaar is ook de vraag wanneer er sprake is van verzamelen van persoonsgegevens. Bij “on the fly” Big Data analysis zal vrijwel altijd sprake van verwerken van persoonsgegevens zijn, het is immers niet goed denkbaar dat er geen

tot de persoon herleidbare informatie uit een dergelijke analyse komt. Als slechts verzameld en opgeslagen wordt, dus een grote hoeveelheid data op een berg gegooid, dan zitten daar ongetwijfeld persoonsgegevens in. Zolang je niks doet met de berg, is de vraag of de berg gegevens als zodanig juridisch relevant is. Over de eventuele inbreuken op de persoonlijke levenssfeer die het enkele vergaren mee kan brengen bij activiteiten van de inlichtingendienst is het rapport Dessens[29] (p. 74-75) in ieder geval lichtvaardig:

> Er wordt hier nog geen kennis genomen van de inhoud van de telecommunicatie waardoor er volgens de wetgever nog geen sprake is van een inbreuk op de persoonlijke levenssfeer, meer in het bijzonder het telefoon- en telegraafgeheim. Van een dergelijke inbreuk is volgens de wetgever pas sprake op het moment dat de gegevens geselecteerd worden.

Telecommunicatiegegevens vertellen steeds meer over ons, waar we zijn, met wie we spreken, etc. Het feit dat iemand in de gaten wordt gehouden en dat dit consequenties voor deze persoon kan hebben in de toekomst, kan een conformerend effect op personen hun handelen hebben dat ook wel omschreven wordt als de disciplinerende werking van toezicht.

Bij nog niet geanalyseerde gegevensverzamelingen bedoeld voor Big Data analysis is er een voortdurende dreiging dat er tot de persoon herleidbare informatie uit gedestilleerd zal worden. Of de informatie gebruikt gaat worden of niet is irrelevant voor de vraag of de Wbp, danwel de Wpol van toepassing is, omdat beide wetten bepalen dat het verzamelen van persoonsgegevens een verwerking betreft. De inlichtingen- en veiligheidsdiensten kunnen nagenoeg onbeperkt gegevens verzamelen. Dit tezamen met de continue toenamen van

[29] Evaluatie Wet op de inlichtingen- en veiligheidsdiensten 2002, Naar een nieuwe balans tussen bevoegdheden en waarborgen..

digitale sporen die burgers achterlaten zorgt voor een niet aflatende inperking van de autonomie van burgers.

4.1.5 Gebrek aan transparantie

De Artikel 29 Werkgroep (WP29),[30] wijst op enkele risico's en uitdagingen inzake het gebruik van Big Data. Zo wordt aangegeven dat transparantie nauwelijks aanwezig is, de overheid meer toezicht kan houden en controle krijgen, alsmede dat de gebruikte data inaccuraat kunnen zijn.

Deze bedenkingen worden in de literatuur bevestigd wanneer over de zogenaamde drie paradoxen van Big Data gesproken wordt: de Transparency, Power en Identity paradox (Richards & King 2013). Hiermee wordt bedoeld dat de verzameling van Big Data onzichtbaar is en dat personen zouden moeten weten welke analyses met hun gegevens gemaakt worden. Daarnaast zijn regels omtrent het gebruik van Big Data niet altijd duidelijk waardoor de bedrijven en de overheid veel macht kunnen uitoefenen. Tenslotte worden analyses veelal gebruikt om 'persoonlijke' voorkeuren aan gebruikers te bieden, zoals gepersonaliseerde advertenties op websites, waardoor men zijn eigen 'identiteit' kan verliezen.

De Artikel 29 Werkgroep stelt dan ook voor dat gebruikers/consumenten toegang hebben tot hun (Big) data profiel. Niet alleen om transparantie te bevorderen, maar ook zodat zij het desgewenst kunnen corrigeren. Dit heeft als het goed is vervolgens een positief effect voor Big Data analysis omdat deze hierdoor meer betrouwbaar en volledig zal zijn. De correctiemogelijkheden kunnen voorkomen dat er incorrecte data worden geanalyseerd, waardoor het adagium "Garbage in, garbage out" voor wat betreft het eerste deel niet langer opgaat. Er moet echter niet vergeten worden dat bij

[30] Dit is een adviesorgaan binnen de Europese Unie die zich met privacy bezighoudt. De naam is ontleend aan het artikel waarbij het orgaan in het leven is geroepen, nl. art. 29 Richtlijn 95/46/EG.

Big Data analysis op grond van correcte data ook onjuiste uitkomsten kunnen worden verkregen. Deze imperfectie kan niet voorkomen worden door het controleren van de invoer, maar wel door betere analyse algoritmes te gebruiken en altijd kritisch naar de uitkomsten te kijken.

4.2 Nieuwe randvoorwaarden profileren

Een term die nieuw is in de voorgestelde privacy verordening is pseudonieme gegevens:

> (2a) 'pseudonymous data' means personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution;

Artikel 20 lid 1 voorstel privacy verordening bepaalt:

> Iedere natuurlijke persoon heeft het recht niet te worden onderworpen aan een maatregel waaraan voor hem rechtsgevolgen zijn verbonden of die hem in aanmerkelijke mate treft en die louter wordt genomen op grond van een geautomatiseerde verwerking die bestemd is om bepaalde aspecten van zijn persoonlijkheid te evalueren of om met name zijn beroepsprestaties, economische situatie, verblijfplaats, gezondheid, persoonlijke voorkeuren, betrouwbaarheid of gedrag te analyseren of te voorspellen.

In een stuk van Europese Raad van 16 januari 2014 is deze bepaling aangescherpt:[31]

[31] Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the

> Every data subject shall have the right not to be subject to a decision based solely on automated processing of data intended to evaluate certain personal aspects relating to a natural person, such as his or her performance at work, economic situation, health, personal preferences, or interests, reliability or behaviour, location or movements and which produces legal effects concerning him or her or severely affects him or her unless such processing is subject to suitable measures to safeguard the data subject's rights and freedom and his or her legitimate interests, such as the rights of the data subject to obtain human intervention on the part of the controller, to express his or her point of view, and to contest the decision, and:
> (a) is necessary for the entering into, or performance of, a contract between the data subject and a data controller and suitable measures to safeguard the data subject's legitimate interests have been adduced, such as the rights of the data subject to obtain human intervention on the part of the controller, to express his or her point of view, and to contest the decision 3; or
> (b) is (…) authorized by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's legitimate interests; or
> (c) is based on the data subject's explicit consent (…).

In hetzelfde document is de volgende definitie opgenomen:

> 'profiling' means any form of automated processing (…) intended to create or use a (…) profile by evaluating personal aspects relating to a natural person, in particular the analysis

---

free movement of such data (General Data Protection Regulation) – Profiling, zie http://www.statewatch.org/news/2014/jan/eu-council-dp-reg-profiling-5344-14.pdf

> and prediction of aspects concerning performance at work, economic situation, health, personal preferences, or interests, reliability or behaviour, location or movements

Wat deze bepalingen in de praktijk voor gevolgen zullen hebben is niet precies te zeggen. Het ziet vooral op geautomatiseerde beslissingen die genomen worden op basis van profielen, de verwerking op zichzelf is niet direct aan nadere regels onderworpen. Er is wel discussie over de vraag wanneer gegevens nu gepseudonimiseerd zijn, of het ontbreken van NAW gegevens bijvoorbeeld hierbij volstaat. Gegevens waar de naam uit verwijderd is, maar nog wel geboortedatum en postcode bevat worden gepseudonimiseerd genoemd maar zijn in de regel te herleiden tot een persoon. Het CBS heeft aangegeven dat zij dergelijke gepseudonimiseerde gegevens uit de gezondheidszorg kunnen koppelen aan de namen van personen, de gegevens worden dan van "van vlees en bloed".[32]

Een bijkomende voorwaarde is dat de organisatie niet de mogelijkheid moet hebben om een uniek kenmerk weer terug te vertalen naar de weggelaten NAW gegevens. Dit zal echter ook tot de mogelijkheden behoren zonder dat er een database is met een tabel waarin een uniek kenmerk aan NAW gegevens gekoppeld. Ook zonder NAW gegevens is het, zeker als er van internet gegevens afgehaald worden of anderszins grote verzamelingen gegevens beschikbaar zijn, veelal mogelijk geanonimiseerde of gepseudonimiseerde gegevens te herleiden tot een persoon.

---

32 http://www.joop.nl/leven/detail/artikel/26491_jacht_op_uw_medische_gegeven_in_volle_gang/

4.3 De relevantie van het EVRM

Het Europees Verdrag voor de Rechten van de Mens tezamen met het Europees Hof voor de Rechten van de Mens, hebben een belangrijke taak als hoeder van de mensenrechten in Europa. Het EVRM werkt door in de Nederlandse rechtsorde en dit betekent dat de overheid ook aan haar inhoud gebonden is. Artikel 8 lid 1 EVRM luidt als volgt:

> Een ieder heeft het recht op respect voor zijn privé leven, zijn familie- en gezinsleven, zijn woning en zijn correspondentie.

Het recht op privacy is de uitgangspositie van het individu in de maatschappij. Een toestand waarin de burger vrij is van willekeurige inmenging door de overheid. Iedere inmenging van de overheid is een uitzondering op deze regel en moet voldoen aan de eisen die daaraan worden gesteld, zoals geformuleerd in het tweede lid van artikel 8 EVRM:

> Geen inmenging van enig openbaar gezag is toegestaan in de uitoefening van dit recht, dan voor zover bij de wet is voorzien en in een democratische samenleving noodzakelijk is in het belang van de nationale veiligheid, de openbare veiligheid of het voorkomen van wanordelijkheden en strafbare feiten, de bescherming van de gezondheid of de goede zeden of voor de bescherming van de rechten en vrijheden van anderen.

Big Data analysis door de overheid moet aldus een wettelijke basis hebben en deze wet moet voor de burger kenbaar en voorzienbaar zijn. Dit houdt in dat de wet toegankelijk en voldoende nauwkeurig moet zijn. Ook moet de wet bescherming bieden tegen willekeurige inmenging door publieke autoriteiten.[33] De meest substantiële

[33] The right to respect for private life handbook, p. 25.

voorwaarde die wordt gesteld aan een inmenging is dat deze noodzakelijk moet zijn in een democratische samenleving en deze eis houdt een proportionaliteits- en subsidiariteitstoets in. Dat wil zeggen dat er in het geval van Big Data analysis per geval moet worden gekeken of de verwerking van persoonsgegevens en de inbreuk op de privacy die daarmee gepaard gaat in verhouding staat tot het te realiseren doel. Hierbij moet worden gekeken naar de aard van de gegevens, het belang van de burger bij het beschermen van zijn of haar gegevens, de aard van de inmenging en welke dwingende maatschappelijke noodzaak er bestaat om deze gegevens te verwerken.

Dit is een tamelijk complexe aangelegenheid die een grote verantwoordelijkheid voor de overheid inhoudt, indien zij toch besluit over te gaan tot een analyse. Artikel 8 EVRM speelt bij de toepassing van de Wbp en de Wpol op de achtergrond een belangrijke rol. Een recent voorbeeld waarbij deze rol duidelijk naar voren kwam, was toen de Belastingdienst aan parkeerbedrijf SMSParking verzocht om de parkeergegevens van al haar klanten over te dragen. Dit was een ongericht verzoek waarbij de Belastingdienst zelf de gegevens wilde filteren op relevantie, waarbij zij van tevoren had aangegeven dat de gegevens zouden worden gebruikt voor niet alleen motorrijtuigenbelasting, maar onder andere ook inkomstenbelasting, loonbelasting, vennootschapsbelasting en omzetbelasting. SMSParking weigerde dit en werd gedaagd voor de voorzieningenrechter in ’s-Hertogenbosch, welke bepaalde dat een dergelijk ongebreidelde opvraag deed vermoeden dat de Belastingdienst de uitzondering van artikel 8 lid 2 EVRM hanteerde als de regel.[34] De overwegingen van de rechter geven duidelijk aan dat het uitwerpen van elektronische sleepnetten om overtredingen van burgers op te sporen, ook wel bekend als *rasterfahndung*, in een democratische samenleving aan banden moet worden gelegd. Hij overwoog het volgende:

[34] Rb. ‘s-Hertogenbosch (vzr.) 26 november 2013, 4.17, ECLI:NL:RBOBR:2013:6553.

> ‘De tot het openbaar gezag gerichte en ter bescherming van de burgers strekkende hoofdregel in artikel 8 EVRM luidt voor zover in dit geval relevant: “een ieder heeft recht op respect voor privéleven” en “geen inmenging van enig openbaar gezag is toegestaan in de uitoefening van dit recht”. Dit uitgangspunt in de relatie tussen burger en overheid is niet het veelgehoorde “wie niets te verbergen heeft, heeft ook niets te vrezen” maar “het dagelijks doen en laten van de burgers gaat de overheid niets aan”.[35]

Het is niet zeker of het combineren van alle informatie uit de SMSParking registratie met bij de belastingdienst aanwezige bestanden, eventueel gecombineerd met op internet aanwezige informatie, als Big Data analysis gekwalificeerd kan worden. Wat deze zaak wel duidelijk maakt is dat in geval van Big Data analysis van de overheid afwegingen mede in het licht van privacy als grondrecht worden verwacht.

[35] Rb. ‘s-Hertogenbosch (vzr.) 26 november 2013, 4.15, ECLI:NL:RBOBR:2013:6553.

5 Big Data gebruik in de rechtspraak

ODR should not be thought of only in the context of Dispute Resolution but possibly also as a dual use space that might generate rich data of interest to national security

Ben Davis, UN ODR Forum Montreal 2013

Op dit moment wordt binnen de Nederlandse rechtspraak nog niks gedaan met Big Data analysis en er lijkt ook nog geen idee te bestaan over hoe Big Data analysis gebruikt zou kunnen worden.[36] Hierin is de rechtspraak niet uniek, want ook veel bedrijven menen iets met Big Data te moeten doen zonder een concrete voorstelling te hebben over op welke wijze (Ross, Beath & Quaadgras 2013). We bespreken hieronder enkele mogelijke toepassingen van Big Data analysis binnen de rechtspraak.

5.1 Management informatie

Big Data analysis zou kunnen worden toegepast bij het genereren van management informatie. Al in de jaren negentig werd voor management informatie gebruik gemaakt van een speciaal daarvoor ontwikkeld programma, namelijk RAPSODY (Lodder & Oskamp 2001). Eind jaren negentig zijn de zogenaamde Lamicie-normen ontwikkeld om de omvang van de werkzaamheden van rechters te standaardiseren en de verdeling van budgetten over de verschillende gerechten op basis daarvan te laten plaatsvinden (Van der Knaap & Van der Broek 2000). Deze normen werden in 2002 ingevoerd, niet zonder toeval ook het jaar dat de Raad voor de Rechtspraak werd ingesteld.[37] Hoe meer informatie elektronisch beschikbaar is, des te

[36] De persoon die wij op aanraden van de Directeur Strategie en Ontwikkeling bij Raad voor de rechtspraak hierover spraken had verwacht dat wij met suggesties over mogelijke toepassingen zouden komen, maar had er zelf nog geen voorstelling bij.

[37] *Bekostiging, doelmatigheid, kwaliteit rechtspraak*, Verslag symposium bekostiging commissie Deetman, Utrecht 2006: "In 2002 werd de eerste versie van het Bekostigingsmodel voor Gerechten ingevoerd".

uitgebreider de mogelijkheden om hieruit voor het beheer en bestuur van de gerechten interessante informatie te verkrijgen. Op dit moment is de verwachting dat in 2018 de rechtspraak volledig elektronisch is (Van Dijk & Van den Hoogen 2014):

> "Binnen vier jaar werken alle rechters in Nederland volledig digitaal, kan iedereen zijn stukken digitaal bij de rechtbank indienen en zijn professionele partijen verplicht om dit te doen."

Door deze digitalisering zal de informatie die gebruikt kan worden voor Big Data analysis ook alle stukken van partijen omvatten en dus niet enkel de uitspraken. Indien er op termijn op grote schaal gegevens geanalyseerd worden, biedt dit mogelijkheden tot uitgebreidere analyses dan bij klassieke management-informatie het geval is.

Leidinggevenden en managers mogen binnen een professionele organisatie, zoals de rechtspraak, gebruik maken van informatie over de in de organisatie werkzame personen, zoals rechters. Aan dit gebruik worden in het algemeen grenzen gesteld. Zo is het continu met camera opnemen van de werkzaamheden van werknemers technisch mogelijk. Juridisch is dit niet aanvaardbaar, tenzij bijvoorbeeld de bedrijfsvoering daar aanleiding toe geeft (een juwelier) of personen verdacht worden van frauduleuze praktijken. Het monitoren van de werkzaamheden van rechters moet met de nodige terughoudendheid worden toegepast, omdat dergelijke activiteiten op gespannen voet staan met hun onafhankelijkheid.

Big Data analysis opent de mogelijkheid om allerhande conclusies te verbinden aan bijvoorbeeld de toetsaanslagen, internetgebruik of de wijze waarop een vonnis geconcipieerd wordt zoals al dan niet gebruik makend van standaardteksten of stukken uit eerdere vonnissen. Aan gedrag gerelateerde activiteiten monitoren en analyseren past goed binnen de mogelijkheden van Big Data

toepassingen. Ook kan de doelmatigheid van de rechtspraak door dergelijke analyses worden verbeterd, dan wel zou de (in)efficiëntie van rechters kunnen worden blootgelegd. Dergelijke analyses kunnen alleen plaatsvinden na overleg met de betrokkenen en wanneer er een duidelijke vraag aan de analyse ten grondslag ligt. Met bestaande management-informatie kan bijvoorbeeld worden vastgesteld dat een rechter bovengemiddeld vaak een specifiek type zaken behandelt of steeds zaken naar zich toetrekt waar een bepaald bedrijf als partij bij betrokken is, zonder dat hier op voorhand een verklaring voor is. De mogelijkheden tot in detail allerhande eigenschappen en kwaliteiten van een rechter bloot te leggen dienen zich aan, maar is enkel toelaatbaar als nagedacht is over het probleem waarvoor de analyse wordt ingezet en het doel dat met de gegevensverwerking beoogd wordt.

5.2 Voorspellen van uitspraken

In de medische sector kunnen patronen worden blootgelegd door Big Data analysis waarmee verband tussen gedrag en ziektes kan worden gelegd. Het is goed denkbaar dat bepaalde uitingen of formuleringen van partijen in positieve of negatieve zin bijdragen aan het vonnis. Deze informatie zou kunnen worden gebruikt om factoren te identificeren en aan te geven op welke wijze deze bijdragen aan de uitkomst van een zaak. Een terrein waar altijd veel belangstelling bestaat voor voorspellingen is de straftoemeting. In de jaren negentig zijn daar verschillende modellen voor ontwikkeld, zowel in Nederland (Eduard Oskamp) als in bijvoorbeeld Schotland (Cyrus Tata) en Israel (Uri Schild).

Het Openbaar Ministerie heeft de zogeheten Polaris richtlijnen opgesteld die gebruikt worden om de strafmaat te bepalen bij kleinere, veel voorkomende delicten.[38] Aan de hand van een aantal factoren (bijvoorbeeld delicttype, mate van toegebracht leed) wordt

[38] http://www.om.nl/organisatie/beleidsregels/bos_polaris/

een strafmaat bepaald. Hiervoor is ook een eenvoudig beslissend ondersteunend systeem ontwikkeld (Lodder 2001) dat na aanvankelijk tegenwerking vanuit het OM sinds eind 2002 publiekelijk beschikbaar is.[39] De rechterlijke macht heeft haar eigen[40] systemen ontwikkeld, waarbij de aandacht anders dan bij het OM vooral gericht was op de ondersteuning bij zwaardere misdrijven. Juist hier is het niet altijd eenvoudig om goed gemotiveerd een straf op te leggen, nu in de regel de motivering gebruikt voor een straf van bijvoorbeeld 12 jaar net zo goed zou kunnen zijn gebruikt voor een straf van 11 jaar.

Het leidt geen twijfel dat Big Data analysis ook binnen de rechtspraak patronen kan blootleggen. De vraag is wederom of dit wenselijk is. In ieder geval zal ook hier goed moeten nagedacht over het doel van de verwerking. Hoewel niet alle rechterlijke uitspraken breed gedragen worden, is een voorafgaande schifting bij Big Data analysis niet nodig. Door grote aantallen zaken te analyseren kunnen waardevolle verbanden worden ontdekt. Het patroon kan bijvoorbeeld zijn dat factor A en B in verband te brengen zijn met strafmaat C. Hoewel op grond van de analyse juist, blijkt hieruit niet of en zo ja op welke wijze een andere factor in een voorliggende of geanalyseerde zaak zou moeten meewegen. Ook kunnen er patronen ontdekt worden die men liever niet blootgelegd ziet. Een kritische analyse van en toelichting op de uitkomsten van de analyse is hier noodzakelijk.

In de jaren negentig werd wel beweerd dat neurale netwerken een goede voorspellende waarde zouden kunnen hebben voor het toekennen van de doodstraf door de huidskleur van de dader en het

[39] http://lodder.cli.vu/rechtszaakbos/

[40] De Databank Consistente Straftoemeting (DCST), het verzamelen van uitspraken om zo een beter beeld te krijgen waar minimaal een straf van 4 jaar op staat, uitzonderlijke gevallen. Echter DCST blijkt niet veel gebruikt te worden omdat het juist over uitzonderlijke gevallen gaat die dus telkens op zichzelf moeten worden beoordeeld. http://www.rechtspraak.nl/Organisatie/Publicaties-En-Brochures/Researchmemoranda/research%20memoranda/2007-RM-afgewogen-straffen.pdf

slachtoffer als input factoren te nemen. Zo bleek, en mogelijk geldt dat nog steeds, dat als een zwarte een blanke vermoordde in de USA de kans vele malen groter was dat deze de doodstaf kreeg dan wanneer een blanke een zwarte had vermoord. Dergelijke informatie is pijnlijk en natuurlijk niet iets dat je anders dan voor een verandering in benadering van dergelijke zaken wil gebruiken.

In Nederland is het niet ondenkbaar dat de sociale achtergrond van partijen en verdachten mogelijk ongewild en onbewust meespelen in de oordelen van de rechter. De vraag is of en hoe gepreciseerd (naar rechter) er behoefte aan dergelijke informatie is en wederom moet hier over de wenselijkheid van dergelijke informatie worden nagedacht. Het is overigens niet ondenkbaar dat de correlaties die door Big Data analyse worden ontdekt door de rechters zelf op prijs worden gesteld. Onbewust handelen en gedrag kan als men daar van bewust wordt gemaakt immers wijzigen.

Behalve binnen het strafrecht is Big Data analysis in ieder ander rechtsgebied mogelijk. De charme van het op een grote hoop gooien van informatie gaat in zekere zin verloren als de informatie van te voren geschift moet worden aan de hand van bijvoorbeeld rechtsgebied. Wellicht is dit niet nodig en is de analyse software in staat om zaken naar rechtsgebied te sorteren. Hoe het zij, een terrein dat buiten het strafrecht interessant kan zijn om te analyseren is het aansprakelijkheidsrecht. Het aansprakelijkheidsrecht kent een algemene zorgvuldigheidsnorm die per geval wordt ingevuld. Als een dergelijk norm geschonden is, kan dit al dan niet leiden tot het vestigen van aansprakelijkheid en in geval van vestiging tot een daaraan gekoppelde vergoeding. Het uitgangspunt van het BEST-project[41] (2005-2010) was om een groot aantal uitspraken te

[41] Een door NWO gesubsidieerd ToKeN-project onder leiding van Arno Lodder en Frank van Harmelen. BEST staat voor BATNA Establishment using Semantic web Technologies. BATNA is een uit het Harvard Negotiation project bekende afkorting die staat voor Best Alternative To a Negotiated Agreement. Het idee was om burgers te informeren over hun kansen in een rechtszaak. Deze informatie kon zowel

analyseren om op grond daarvan voorspellingen te kunnen doen bij een nieuwe aansprakelijkheidszaak (Uijttenbroek *et al.* 2008). Op het moment dat het project liep waren er onvoldoende zaken beschikbaar voor een analyse op grote schaal. Hierin zal met de aangekondigde digitalisering van de rechtspraak zeker verandering komen. Het voorlichten van procespartijen over hun kansen kan een belangrijke rol spelen bij het verlichten van de werkdruk van de gerechten. Dergelijke informatie is mogelijk op verschillende terreinen met behulp van Big Data analysis te verkrijgen. Gedacht kan worden aan ontslagzaken, huurconflicten en auteursrechtelijke gedingen.

In Amerika is door een jurist in samenwerking met IT-ers het Big Data analysis programma *Juristat* ontwikkeld. In 2012 begon de voormalige advocaat Andrew Winship samen met Jordan Woerndle en de software ontwikkelaar Bob Ward aan wat eind 2013 een succesvolle toepassing is geworden (Nicklaus 2013). Het programma werkt in het domein van octrooien en verzamelt data van verschillende bronnen waaronder rechtspraak, sociale media, publieke databases en demographics. Vervolgens wordt deze data geanalyseerd en vertaald in nuttige informatie voor advocaten, of zoals hun slogan op de website luidt: “Helping patent lawyers predict the future”. Zo worden voorspellingen gedaan over welke richting de rechter hoogstwaarschijnlijk zal volgen in bepaalde situaties. Dit lijkt op de hierboven beschreven doelstelling van het BEST-project en daaraan gerelateerde mogelijkheden. Veel verder strekt de door het Amerikaanse programma gedane suggesties als op welke dag het gunstig is een bepaalde zaak te laten behandelen door een specifieke rechter. Het idee is dat advocaten op deze manier meer kans maken een zaak te winnen. Het programma informeert volgens de oprichter Andrew Winship een advocaat bijvoorbeeld als volgt (Cherry 2013):

---

gebruikt worden om te beslissen een procedure te beginnen, dan wel om bij onderhandelingen buiten de rechter om een richtpunt te hebben.

> "Hey, there's a case you're involved in with three standard deviations of difference in behavior between the judge you have and the judge you probably want. Here's something that should be in front of your eyeballs that you weren't thinking to look at."

Het succes van een dergelijke programma is ondermeer afhankelijk van de hoeveelheid informatie die de rechtbanken opslaan. Daarnaast is het ook van belang hoe toegankelijk deze informatie is. De hierboven aangegeven terughoudendheid ten aanzien van de analyse van informatie rond rechterlijke uitspraken zal voor bedrijven die zich hier op richten niet opgaan. De verbanden die blootgelegd worden zouden bezwarend of compromitterend kunnen zijn. De rechtspraak zal zich op het moment dat steeds meer informatie vrijelijk beschikbaar komt moeten beraden hoe hiermee om te gaan.

5.3 Gebruik in rechtszaken

Binnen het strafrecht vormt het leerstuk inzake rechtmatig verkregen bewijs de garantie dat niet op alle mogelijke manieren bewijs tegen een verdachte verzameld mag worden. Het civiele recht kent een vrij bewijsstelsel, maar ook daar zou de vraag zich kunnen voordoen of een partij gebruik kan maken van de uitkomsten van Big Data analysis. Het controleren van op deze wijze verkregen uitkomsten is voor een doorsnee rechter niet gemakkelijk. Bovendien is ook hier weer de vraag in hoeverre het door Big Data analysis aantonen *dat* iets zo is, volstaat als niet duidelijk is *waarom* deze uit de berg data volgende conclusie aanvaardbaar is.

De kwestie of statische methoden een rol kunnen spelen in de bewijsvoering is al langer onderwerp van discussie, zo geeft Nissan (2013) aan:

> "Information technologists coming into the field of modelling of the reasoning on legal evidence are all too easily

> prone to embrace statistical techniques, often unaware of the fierce controversy among legal evidence theorists, concerning whether probabilistic methods should be allowed as a metric in guilty or not guilty determinations"

Big Data analysis is geen klassieke statische methode om uitkomsten met een bepaalde zekerheid te voorspellen, maar ook bij Big Data analysis geldt natuurlijk dat 100% zekerheid niet verkregen kan worden. Dit is overigens iets wat zowel in het civiele als het strafrecht geaccepteerd wordt. Iets volledig zeker weten kun je nooit. Daarom wordt in het strafrecht als maatstaf genomen dat zaken wettig en overtuigend bewezen moeten worden, en in het civiele recht moet een bewering aannemelijk zijn. Bij de lichtere civiele maatstaf zal een rechter eerder geneigd zijn uitkomsten van Big Data analyse als bewijs te accepteren.

Over het gebruik van Big Data in rechtszaken zijn ons op dit moment weinig voorbeelden bekend, zeker niet in Nederland. Een Amerikaanse zaak waar Big Data werd gebruikt betreft een man van wie zijn 'bewegingen' via GPS systeem door de politie een maand lang gevolgd werd (*U.S. v. Maynard*, 615 F.3d 544, 555 (D.C. Cir. 2010)). De vraag is of dergelijke informatie als bewijs in een strafzaak kan worden toegelaten of dat het hier onrechtmatig verkregen bewijs betreft. De gevolgde persoon zei dat er sprake was van richtingloos en zonder reden zoeken naar bewijs.

De politie meende echter dat een persoon die zich op openbare grond begeeft in redelijkheid geen privacy verwachtingen[42] kan hebben. Er werd uiteindelijk geconcludeerd dat een persoon weliswaar geen privacy verwachtingen heeft bij de individuele gevolgde bewegingen, maar dat alle bewegingen in zijn geheel en dus het totale beeld daarvan dat een patroon kan vormen verder strekt

[42] "Reasonable expectation of privacy" is het concept dat in Amerika centraal staat bij de beoordeling of er in een bepaald geval sprake van een inbreuk is.

dan de verwachting die iemand ten aanzien van zijn privacy heeft. Dit is een mooie illustratie van het gegeven dat bij Big Data analysis in de regel het geheel meer is dan de som van haar delen. In Nederland geldt op zich dat zonder verdenking niet personen op deze wijze gevolgd mogen worden. Bovendien is bij stelselmatig volgen, waar hier duidelijk sprake van was, toestemming van de rechter-commissaris vereist.

5.4 Big Data (security) en anonimisering

Mede in het kader van de belangstelling die veiligheidsdiensten stellen in via internet beschikbare informatie, verdient het gebruik van online beschikbare vonnissen en andere procesinformatie aandacht. Dit werd door Ben Davis naar voren gebracht in een presentatie *Big Data Security and Online Dispute Resolution* tijdens het 12de UN ODR forum gehouden te Montreal.[43] Zijn observaties werden gedaan in de context van Online (alternatieve) Dispute Resolution, maar kunnen ook doorgetrokken worden naar rechterlijke uitspraken:

> "No lawyer can know the extent to which Big Data Security is operating in every facet of his/ her digital work life or his/her firms "secure" or "encrypted" spaces. Every lawyer can know that Big Data Security is interested in everything. Disputes and dispute resolution would appear to be a particularly interesting area to examine as it shows connections between people."

Enerzijds kunnen surveillance activiteiten van met name veiligheidsdiensten raken aan het beroepsgeheim van advocaten alsmede toegang tot stukken betreffen waarvan men dit liever niet

43 http://www.laboratoiredecyberjustice.org/Content/documents/odr2013/Davis%20Panel%207.pdf

heeft. Uit openbaar beschikbare informatie kunnen zoals Davis aangeeft interessante verbanden tussen personen worden gedestilleerd. Een punt dat hier aandacht verdient is of het doel dat ten grondslag ligt aan het anonimiseren van vonnissen door Big Data toepassingen mogelijk ondermijnd wordt. We gaan er vanuit dat het matchen van uitspraken aan bijvoorbeeld elders op internet aanwezige informatie steeds meer gevallen tot de-anonimisering zal leiden. Dit is op dit moment handmatig al mogelijk door bijvoorbeeld blogberichten en informatie uit de media te combineren met geanonimiseerde vonnissen, maar door Big Data analysis zal dit op grotere schaal met gebruik van veel informatie als vanzelf tot meer treffers leiden. De anonimiteit kan dan gemakkelijker teniet gedaan worden.

6 Big Data gebruik in de opsporing

And I remember my lawyer telling me, "If I was the judge, I'd find you guilty!" You know? I don't know for what. For being, just for being there, I guess
Bruce Springsteen, Growin' up, 7 juli 1978, The Roxy Theatre

Zeker na de onthullingen van Edward Snowden in het voorjaar van 2013 is de verbinding tussen Big Data analysis en activiteiten van inlichtingendiensten volop in de belangstelling gekomen. Vanwege de verschillende taakstelling en het daarbij behorende wettelijk kader moeten inlichtingendiensten wel uitdrukkelijk onderscheiden worden van opsporingsinstanties. Eerst genoemden richten zich op het constateren van dreigingen voor de staatsveiligheid en als die zich voordoen proberen ze hiernaar te handelen. Opsporingsdiensten hebben de taak criminaliteit op te sporen. Er is een snijvlak, aangezien dreigingen voor de staatveiligheid veelal samenhangen met binnen het strafrecht gesanctioneerde activiteiten. Hoewel de focus duidelijk verschilt van vrij abstracte dreigingen (veiligheidsdiensten) tot concrete verdenkingen (opsporing), wordt niettemin informatie tussen beide instanties gedeeld.

Het gebruik van Big Data binnen de Nederlandse opsporing vindt al enige tijd plaats. Schepers (2012) schrijft dat de politie Amsterdam-Amstelland al in 2001 met een Big Data project[44] begon omdat de politie worstelde

> "...met problemen zoals het inconsistent (en handmatig) verzamelen en analyseren van data, lastig ontsluitbare systemen, vervuilde gegevens en verschillende tools met elk hun eigen beperkingen."

De succesvolle implementatie van het systeem vond plaats in 2006 en inmiddels heeft het project ertoe geleid dat het analysewerk

[44] Uiteraard werd de term Big Data in die tijd nog niet gehanteerd.

efficiënter is geworden. Los van de vraag of het project als Big Data project kan worden gezien, heeft het in ieder geval gemeen dat door het project gerichter schaarse middelen worden ingezet en daardoor effectiever te werk wordt gegaan. De nadruk in dit hoofdstuk ligt bij een toepassing, *predictive policing*, die inspeelt op een behoefte naar efficiënte en het zo effectief mogelijk inzetten van schaarse middelen. Daarnaast gaan we in op zogenaamde webcrawlers, toepassingen die het internet doorzoeken naar voor de politie interessante informatie.
Al in 2004 is bij de wijziging van de Wet op de inlichtingen veiligheidsdiensten (Wiv) ingegaan op de mogelijkheid om op grote schaal data te analyseren:

> "(…) het zo vroeg mogelijk identificeren van de voorbereiding van mogelijke terroristische acties en de daders daarvan. Nieuwe vormen van geautomatiseerde data-analyse worden daarvoor ingezet, zoals (…) data-mining. Daartoe moeten grote bestanden met persoonsgegevens van niet-verdachte personen doorzocht worden, hetgeen op gespannen voet kan komen met de thans geldende wettekst (…) De wet zal (…) meer armslag moeten bieden."[45]

Data mining wordt genoemd, maar omdat pas rond die tijd de eerste Big Data toepassingen werden ontwikkeld en de term zelfs nog later in zwang is gekomen werd nog niet verwezen naar Big Data analysis. Drie jaar later is een vergelijkbaar geluid te beluisteren bij de onderzoekscommissie Bosma in het Rapport van de Adviescommissie Informatiestromen Veiligheid:[46]

> "De enorme groei van databanken en communicatiemogelijkheden en de opkomst van geavanceerde technologie om te zoeken in deze grote

[45] *Kamerstukken* II, 2003/04, 29 200 VII, nr 61.
[46] *Data voor daadkracht. Gegevensbestanden voor veiligheid: observaties en analyse*, 2007, http://bit.ly/1fbM0KQ.

hoeveelheid gegevens bieden de inlichtingen- en opsporingsdiensten veel nieuwe mogelijkheden om hun doelstellingen te realiseren."

Ook hier gaat het nog om klassieke analysetechnieken, immers het zoeken in gestructureerde in databanken opgeslagen gegevens.

6.1 Project X en Bestwelsnel.nl

6.1.1 Dreigingsanalyse

In maart 2013 presenteerde de Commissie Cohen het rapport dat naar aanleiding van de Facebook-rellen in Haren is opgesteld. Het ging hier om een meisje uit Haren dat op haar Facebook-pagina een uitnodiging voor haar verjaardag plaatste en abusievelijk deze uitnodiging die enkel bedoeld was voor haar vrienden (privé setting) voor een ieder toegankelijk had gemaakt (openbaar). De uitnodiging ging vervolgens "viral", als een lopend vuurtje werd de uitnodiging gedeeld waardoor al snel duizenden mensen van het feestje op de hoogte waren. De term project X is ontleend aan de gelijknamige film uit 2012 waar een feestje door enorme toestroom van mensen volledig uit de hand loopt.

Deelrapport 2 (Van Dijk e.a., 2013) gaat over de rol van jongeren, sociale media, massamedia en autoriteiten bij de mobilisatie voor Project X Haren. De onderzoekers maakten onder andere gebruik van ruim 52.000 Facebook-berichten die jongeren op de evenement-pagina "PROJECT X – HAREN" plaatsten in de periode van 7 september tot en met 3 oktober 2012. Door gebruikers geüploade afbeeldingen zijn niet opgenomen in de dataset. Wel zijn alle verwijzingen vanuit berichten met een *url* naar externe websites, video's en afbeeldingen in de dataset opgenomen.

Ook is gekeken naar Twitter data, dit betreft ruim 500.000 berichten. Door deze data achteraf te analyseren, meende de commissie Cohen

beter te kunnen verklaren hoe de situatie in Haren kon ontstaan, bijvoorbeeld door het netwerk van contacten van bij de rellen betrokkenen te bestuderen. Zo werd getoond dat er een kern was van ongeveer 250 gebruikers die de *hype* hebben opgestart en in stand hielden, en dat deze kern hoofdzakelijk uit Haren of de directe omgeving kwam. Of we met het *real-time* analyseren van dergelijke data het Haren-incident hadden kunnen voorkomen is een zeer interessante kwestie die zich leent voor nader onderzoek.

Hierbij moet wel bedacht worden dat wanneer teruggekeken wordt vanuit een incident gericht analyseren mogelijk is. Bij beschikbaarheid van grote hoeveelheden data zonder een al bekend richtpunt, kan er wel Big Data analysis worden toegepast maar is de kans op succes uiteraard kleiner dan wanneer er op voorhand al duidelijke richtpunten zijn. Op het moment dat er bijvoorbeeld een aanslag plaatvindt, is het redelijk eenvoudig van daaruit terug te kijken en berichten aan te wijzen waaruit voornemens blijken. Voorafgaand aan een incident zijn er waarschijnlijk tientallen, zo niet honderden signalen die mogelijk tot incidenten kunnen leiden en waarvan er uiteindelijk maar één plaatsvindt.

Het is de vraag of Big Data analysis een oplossing kan bieden om zodanig te differentiëren dat alleen de datastromen die daadwerkelijk leiden tot incidenten in kaart worden gebracht. Uiteraard is optimalisatie mogelijk, maar de juiste balans vinden zal niet gemakkelijk zijn. De analyse moet uiteindelijk leiden tot zoveel mogelijk terechte signalen en zo weinig mogelijk false positives. Dergelijke afwegingen worden in de wereld van de information retrieval al tientallen jaren gemaakt, door een balans te zoeken tussen recall (zoveel mogelijk van hetgeen naar gezocht wordt vinden) en precision (naar verhouding zoveel mogelijk juiste treffers binnen de gevonden resultaten). Een geval als Haren is in die zin gemakkelijker dan bommeldingen via Twitter. Al ruim van te voren was bekend dat er veel belangstelling voor het feest was, dus de analyse van de harde kern had on the fly kunnen plaatsvinden. Deze

informatie had gebruikt kunnen worden om de hype tijdig te neutraliseren.

Bij bommeldingen via Twitter is het lastig om vast te stellen of een scholier nu echt van plan is een bom te plaatsen bij de school. Hoewel in de meeste gevallen dit niet zo zal zijn, pakt de politie regelmatig scholieren op. Begrijpelijk, want het hoeft maar een keer mis te gaan en de publieke opinie zal dan zijn dat de politie het had kunnen weten. Hoewel lastig, kan Big Data analysis om in termen van information retrieval te spreken de recall en precision verbeteren. Alle bommeldingen uit een stroom tweets halen is vrij eenvoudig, de kunst is te kunnen schiften tussen echte dreigingen en niet gemeende dreigingen. Hier is niet mee gezegd dat iedereen maar alles mag roepen op Twitter, maar als er keuzes gemaakt moeten worden over welke scholier opgepakt wordt kan Big Data analysis daar mogelijk bij helpen.

6.1.2 Alternatieve snelheidsmeting

De website Bestwelsnel.nl is een toepassing die door een particulier is ontwikkeld. Hier wordt gebruik gemaakt van sensorinformatie zoals die door lussen onder het wegdek verzameld wordt. Deze informatie is sinds de tweede helft van 2013 voor een ieder beschikbaar en de beheerder van de site Bestwelsnel.nl. gebruikt deze lusinformatie om topsnelheden om Nederlandse snelwegen te berekenen. Lussen op de snelweg liggen namelijk op vaste afstanden van elkaar. Als er op een bepaald tijdstip, doorgaans ’s nachts, maar één voertuig op een weg rijdt kan daarvan de snelheid bepaald worden. Op deze site wordt op de kaart van Nederland aangeven op welke (snel)weg het dagrecord en verder alle metingen boven de 170 km/h. Dit zijn er iedere dag verscheidene.

Interessant is de vraag of iemand beboet kan worden als een agent deze midden in de nacht bij een afrit opwacht nadat hij via Bestwelsnel.nl constateerde dat deze veel te hard reed. Mede vanwege de mogelijkheid op fouten bij Bestwelsnel.nl denken wij niet

dat een rechter hierin in mee zal gaan, maar dit wordt anders als de politie zelf deze informatie zou verwerken. Wellicht een toepassing voor de toekomst. Uiteraard zijn er nu al camera's die gebruikmaken van de lusinformatie, maar mogelijk dat ook op plekken waar er geen camera's zijn deze informatie dus gebruikt zou kunnen worden op de manier waarop Bestwelsnel.nl werkt.

Een ander mooi voorbeeld van alternatieve snelheidsmeting is dat de politie in België van plan is om alle flitscamera's te verwijderen en in plaats daarvan sensors met RFID-tags in de auto's plaatsen die de snelheid meten. Aan het einde van het jaar wordt dan gewoon een rekening gestuurd naar automobilisten.[47]

Lusinformatie zou eventueel ook gebruikt worden als een verdachte beweert ergens gereden te hebben en de lusinformatie aantoont dat er op dat tijdstip niemand was. Big Data analysis van lusinformatie kan zeker meer opleveren dan enkel het in kaart brengen en beheersen van verkeersstromen, zeker als deze informatie gecombineerd wordt met andere informatie.

6.2 Predictive policing

Bij predictive policing wordt informatie uit het verleden gebruikt om voorspellingen te doen die de opsporing ondersteunen. Hoewel de toelaatbaarheid van predictive policing niet echt een punt van discussie is, staan we kort stil bij de bevoegdheid die opsporingsinstanties, met name de politie heeft. De algemene politietaak is het handhaven van de openbare orde (art. 3 Politiewet). De toepassingen die binnen predictive policing worden gebruikt vallen daaronder. Voor het zoeken op internet volstaat deze algemene instructienorm niet onder alle omstandigheden. We komen hier bij de behandeling van webcrawlers op terug. Een

[47] http://www.frankwatching.com/archive/2013/05/17/big-data-hoe-pas-je-het-toe/

belangrijke uitspraak voor de randvoorwaarden van het inzetten van technologie in verhouding tot privacy is het Zwolsman-arrest:[48]

> Uitoefening van de bevoegdheden dient in verhouding tot het beoogde doel redelijk en gematigd te zijn.: de voortschrijdende ontwikkeling van het fundamentele recht op bescherming van de persoonlijke levenssfeer en de toenemende technische verfijning en intensivering van onderzoeksmethoden en -technieken verlangen een meer precieze legitimatie in de wet (r.o. 6.4.4)

Bij predictive policing lijken in beginsel de bestaande bevoegdheden te volstaan. Met het oog op Big Data analysis moet nog genoemd worden artikel 11 van de Wet Politiegegevens:

> Voor zover dat noodzakelijk is voor een onderzoek als bedoeld in artikel 9, eerste lid, kunnen politiegegevens die voor dat onderzoek zijn verwerkt, geautomatiseerd worden vergeleken met andere politiegegevens die worden verwerkt op grond van artikel 8 of 9 teneinde vast te stellen of verbanden bestaan tussen de betreffende gegevens.

De noodzakelijkheid kan bij Big Data analysis in de regel wel hard gemaakt worden. Lastiger is de vraag of de bij Big Data analysis gebruikte gegevens kwalificeren als politiegegevens (art 1 Wet Politiegegevens): " elk persoonsgegeven dat in het kader van de uitoefening van de politietaak wordt verwerkt."

Dit is een algemene definitie waar alle gegevens onder vallen die nodig zijn voor het goed functioneren van de politie, maar het vergaren van grote gegevensverzamelingen of het real-time analyseren van op het internet aanwezige gegevens kunnen alleen

[48] HR 19 december 1995, *NJ* 1996, 249.

dan als politiegegevens worden beschouwd als er een aanwijsbare reden is waarom gezocht wordt.

De bevoegdheden van inlichtingendiensten zijn veel ruimer, zo is in Artikel 6 lid 2 Wiv te lezen:

> De Algemene Inlichtingen- en Veiligheidsdienst is slechts bevoegd gegevens te verzamelen ten behoeve van het opstellen van de dreigings- en risicoanalyses, bedoeld in artikel 6, tweede lid, onderdeel e, indien de gegevens die op grond van het eerste lid zijn verstrekt dat noodzakelijk maken.

Het zal doorgaans niet lastig zijn om noodzakelijkheid aan te tonen dan wel te construeren. Verder kunnen wat de AIVD betreft nog bijzondere bevoegdheden worden genoemd zoals gerichte interceptie (artikel 25 Wiv), selectie na ongerichte interceptie (artikel 27 Wiv) en searchen (artikel 26 Wiv).

Door op grote schaal data te verwerken kunnen patronen bloot gelegd worden die met uitsluitend menselijke capaciteiten lang niet allemaal herkend zouden worden. Deze patronen kunnen bijvoorbeeld belangrijke aanwijzingen geven voor “potentiële hotspots voor criminaliteit” (Doorenbosch 2013). Dit heeft geleid tot de introductie van predictive policing, een werkwijze die voortbouwt op intelligence led policing (Bachner 2013). Intelligence led policing, ook wel bekend als informatiegestuurde politiezorg, heeft als uitgangspunt dat politieacties op basis van inlichtingen plaats vinden. Predictive policing breidt dit idee uit door op grotere schaal informatie te verzamelen, te verwerken en te analyseren. Volgens Perry *et al.* (2013) staat predictive policing voor

> “the use of analytical techniques to identify promising targets for police intervention with the goal of preventing crime, solving past crimes, and identifying potential offenders and

> victims."

Perry *et al.* (2013) identificeren dus vier brede categorieën van predictive methoden: voor het voorspellen van criminele activiteiten, voor het voorspellen van mogelijke daders, voor het voorspellen van mogelijke slachtoffers en voor het identificeren van de identiteit van daders van reeds gepleegde misdaden. Deze laatste categorie is merkwaardig, want het voorspellende element ontbreekt in het geheel op het moment dat een misdrijf al heeft plaatsgevonden.

Idealiter zou met behulp van predictive policing op basis van met Big Data analysis ontdekte patronen kunnen worden geanticipeerd op misdaden, daders en toekomstige slachtoffers. Dit is ook zoals Mayer-Schonberger & Cukier (2013, p. 158) het zien:

> "predictive policing: using big-data analysis to select what streets, groups, and individuals to subject to extra scrutiny, simply because an algorithm pointed to them as more likely to commit a crime"

Hierdoor zouden de analytische mogelijkheden van Big Data aan kunnen sluiten op de gelegenheidstheorie (Felson & Cohen 1980; Felson 1986) zoals deze gehanteerd wordt binnen de criminologie. De informatie die ingewonnen kan worden via predictive policing kan namelijk vervolgens gebruikt worden als input voor het ontwikkelen van een barrière model. Zoals Mitchell (2013) schrijft, predictive policing

> "...sounds like a page from the script of the movie *Minority Report*, but the objective isn't to arrest people but to anticipate and remove the opportunity to commit crimes before they can occur."

Mitchell benadrukt het belang van het reduceren van 'opportunities'. Er zijn ook commerciële toepassingen, zoals ontwikkeld door het

bedrijf PredPol, kort voor predictive policing. Ten aanzien van de vraag hoe ze hun analyses uitvoeren, schrijven ze het volgende:

> "Predictive Policing or PredPol applies advanced mathematics and adaptive computer learning in contrast to technology which simply maps crime data."

Op basis van ingevoerde data, waaronder datum, tijdstip, type delict en locatie, wordt middels een algoritme berekend op welke plek en op welk tijdstip de kans op een bepaalde misdaad, zoals een overval of een inbraak, hoog is. Dit algoritme is ontwikkeld door onderzoekers aan de Universiteit van Californië, Los Angeles en Santa Clara University. Ten grondslag aan het algoritme ligt antropologisch en criminologisch onderzoek (Friend 2013).

Het systeem dat gebruikt wordt door het politiekorps van de stad Santa Cruz in Californië heeft tussen de 1200 en 2000 data punten nodig om het meest betrouwbare beeld te geven. Aangezien Santa Cruz, bijvoorbeeld, tussen de 400 en de 600 inbraken per jaar heeft, gebruikt de politie data van de afgelopen vijf jaar als invoer in het systeem (Friend 2013). Het systeem werkt ook voor andere misdaden, zoals geweld tussen bendes, diefstal, en drugsmisdaden. Voor moordaanslagen is predictive policing echter minder bruikbaar omdat de hoeveelheid data punten te beperkt is om een betrouwbaar beeld te geven. Mogelijk dat in de toekomst de analyse uitgebreid kan worden met grotere hoeveelheden, ongestructureerde data afkomstig van camerabeelden, internet, GPS-informatie, etc. wat nog nauwkeuriger voorspellingen kan opleveren dan met de huidige toepassingen mogelijk is.

Volgens Greengard (2012) is het proces van predictive policing vrij recht toe recht aan. De politie voert data over criminele activiteiten, inclusief datum, tijdstip, locatie en type delict, in het programma en op basis daarvan kan de software 'voorspellingen' geven. Desondanks is het vervolgens noodzakelijk voor opsporingsinstanties om te

begrijpen waarom de kans op een criminele activiteit in een bepaalde regio en op een bepaald tijdstip hoog is. Die kennis is immers een vereiste om vervolgens actie te ondernemen en de kans in de toekomst te verkleinen, oftewel de gelegenheid tot het plegen van een misdaad te reduceren. Deze constatering is belangrijk om te maken, mede omdat predictive policing, ondanks de grote belofte, ook vatbaar is voor een aantal misvattingen. Volgens Perry et al. (2013) is de mens in het gehele proces van predictive policing nog steeds de belangrijkste schakel. Hoewel software dus veelal gepromote wordt als een 'end-to-end solution' is het aan opsporingsagenten zelf om de relevante data te vinden, te verzamelen en ze te bewerken zodat ze geanalyseerd kunnen worden.

Een zorg, geuit door Ferguson (2012), is de invloed van predictive policing methoden op redelijkerwijze verdenking van een verdachte. Zoals Ferguson (2012) zich afvraagt:

> "[a]re data-driven 'hunches' any more reliable than personal 'hunches' traditionally deemed insufficient to justify reasonable suspicion?"

De verantwoording voor een verdenking kan veranderen door het gebruik van Big Data en in het verlengde daarvan predictive policing. Indien niet naar bepaalde plekken, maar naar specifieke personen gekeken wordt is de vraag of dit enkel op grond van een Big Data verdenking toelaatbaar is. Is de verdenking als uitkomst van Big Data analysis een acceptabele grondslag voor een search warrant of een arrestatiebevel? Dat is deels afhankelijk van de betrouwbaarheid van de analyse. Ferguson waarschuwt voor een blind geloof in de voorspellende waarde van predictive policing. Daarom geeft Ferguson aan dat het belangrijk is om te begrijpen waarom de 'voorspellingen' werken, oftewel na te denken over de logica achter een verdenking. Ferguson schrijft dat het gebruikte algoritme geen 'magic box' is, maar probability modellen genereert voor toekomstige

gebeurtenissen op basis van huidige en historische omgevingskwetsbaarheden. Wellicht nog belangrijker is dat Ferguson beargumenteert dat de waarschijnlijkheid dat een misdaad gepleegd gaat worden niet wordt geconstateerd omdat er eerder een misdaad gepleegd is, maar omdat de omgevingsfactoren die de vorige misdaad hebben gefaciliteerd nog steeds onbeantwoord zijn gebleven. Dus, stelt Ferguson, als de kwetsbaarheid verholpen is dan zou de voorspelling irrelevant moeten zijn. Verder benadrukt Ferguson dat de verzamelde data die ingevoerd wordt getoetst moeten worden op betrouwbaarheid, nauwkeurigheid en transparantie. Als predictive policing ten grondslag ligt aan een verdenking, dan zal inzage gegeven moeten worden in de manier waarop tot die voorspelling gekomen is. Dat zal specifieker moeten dan uitsluitend te zeggen dat het op basis van data en een algoritme is, hoe complex het ook moge zijn.

In het verlengde van het gebruik van predictive policing als grondslag voor een verdenking, is een van de meest in het oogspringende valkuilen, tevens besproken door Perry *et al.* (2013), de kwetsbaarheid om bepaalde burgerrechten onvoldoende in acht te nemen. Überhaupt de gedachte, zo schrijven Perry *et al.* (2013), dat bepaalde regio's wel of niet aandacht van opsporingsinstanties verdienen is discutabel vanuit een burgerrechten perspectief. Deze veronderstelling is verrassend aangezien dit 'labelen' zoals zij het noemen ook al voor het gebruik van predictive policing plaatsvond en vanuit een risico management perspectief goed te verklaren is. Bepaalde gebieden, zoals uitgaansgelegenheden, zijn immers risicovoller dan andere gebieden. In de Verenigde Staten is er echter wel degelijk een burgerrechten verband. Daar heeft het hoogste gerechtshof, de United States Supreme Court, gesteld dat de maatstaven voor een redelijke verdenking minder streng toegepast hoeven te worden in 'high-crime areas.' Hierdoor worden de Fourth Amendment rechten van burgers mogelijk beperkt als ze in een high-crime area wonen; hetgeen het labelen van een bepaald gebied als 'high crime' ook politiek gevoelig maakt.

Naïef is de stelling van Perry et al. (2013) dat het identificeren van 'hot spots' en het gebruiken van predictive policing niet per se op gespannen voet staan met privacy rechten omdat de verzamelde informatie geen persoonlijk identificeerbare informatie bevat. Zoals eerder beschreven vereist het algoritme uitsluitend het tijdstip, de datum, de locatie en het type delict. Dit wordt vervolgens vergeleken met historische criminaliteit data en omgevingsfactoren om tot de voorspellingen te komen. Hier komen dus geen 'persoonsgegevens' aan te pas. Hoewel op zich zelf juist, vindt de mogelijke privacy inbreuk niet plaats door de invoer, maar de uitvoer. De conclusies die op grond van predictive policing getrokken worden kunnen wel degelijk van invloed zijn op de persoonlijke levenssfeer van degene die nadeel van deze conclusies ondervinden door bijvoorbeeld in de gaten of aangehouden te worden. Andrew Murray noemde tijdens het congres BILETA 2014 het voorbeeld van een Amerikaans politiecorps dat door Big Data analysis de top 100 potentiële criminelen had geïdentificeerd. Dergelijk personen beter in de gaten houden ligt dan voor de hand. Dat de politie vervolgens bij deze 100 personen aanbelde om te melden dat ze nauwlettend in de gaten gehouden worden, is een te verstrekkende inbreuk op de privacy. Tussen die 100 mensen zullen immers als vanzelf ook personen zitten die geen criminele intenties hebben en ook nooit zullen krijgen.

Hier moet ook gewaakt worden voor goedkope retoriek. Zo werd begin 2014 op een marketing congres gesteld dat als nu elke terrorist opgespoord kan worden door ongebreidelde Big Data analysis, er daardoor nooit meer op het vliegveld gecontroleerd hoeft te worden en het enige nadeel is dat af en toe een onschuldige enkele uren op een vliegveld ten onrechte wordt vastgehouden. Wie zou daar tegen zijn? Behalve dat er mensen om principiële redenen tegen ongebreidelde Big Data analysis zullen zijn, is hier natuurlijk vooral een probleem dat nooit alle terroristen opgepakt kunnen worden. Daarnaast zullen die enkelingen die ten onrechte opgepakt worden er vast meer zijn dan gedacht en ze mogelijk ook langer worden

vastgehouden. Zoals wel vaker bij technologie moet het worden ingezet als hulpmiddel, niet als een panacee.

6.3 Internetopsporing en webcrawlers

iColumbo is een near real time internet monitoring service die door geautoriseerde gebruikers binnen de overheid kan worden gebruikt. iColumbo is zo ontwikkeld dat de technologie voldoet aan forensisch technische eisen waardoor het geschikt is voor opsporing en bewijsvoering. Door het gebruik van open source sofware en open standaarden is de technologie mede vanuit het perspectief van privacy toetsbaar.

Webvoyager is een webcrawler. Crawlers werken vergelijkbaar met een index-robot van Google, via het volgen van links worden grote delen van het internet bekeken en geïndexeerd. Deze toepassingen worden ook wel webspiders genoemd.[49] Webvoyager zoekt tijdens het "afstruinen" van het internet naar verdacht materiaal. Er wordt hierbij niet alleen tekst geanalyseerd, maar ook foto's en beeldmateriaal wordt in de analyse betrokken. Hierbij wordt het programma Impala van EUvision gebruikt om beeldmateriaal te analyseren. Sinds enige tijd ook een App op de markt heeft:[50]

> "Impala is the first app in the world that automatically sorts the photos on your phone. (…) Impala "looks" into your images and videos and recognizes what they are about."

Impala van EUvision moet niet verward worden met Cloudere Impala. Blijkbaar is de term Impala populair in de wereld van Big Data

[49] Voor technische en met name juridische achtergronden van webcrawlers, zie Boonk & Lodder (2006).
[50] http://www.euvt.eu/

analyse, want Cloudera Impala combineert klassieke data analyses met Big Data analyse:[51]

> "Cloudera Impala is the industry's leading massively parallel processing (MPP) SQL query engine that runs natively in Apache Hadoop. The Apache-licensed, open source Impala project combines modern, scalable parallel database technology with the power of Hadoop."

Na de vertrouwelijke eindrapportage over Webvoyager in maart 2013 zijn de ontwikkelingen mede door de herindeling van de politie stil komen te liggen. Begin 2014 volgde het bericht[52] dat het vervolg van Webvoyager niet helemaal duidelijk was, maar de verwachting was dat het doorgezet zou worden. Bij de auteurs is de huidige status van het project niet bekend.

Het analyseren van informatie kan door grote hoeveelheden data binnen te halen en lokaal te analyseren. Dit is wat voornamelijk op dit moment gebeurt. Een andere mogelijkheid is om de gegevens direct van internet te analyseren zonder ze binnen te halen. Dit heeft als voordeel dat er lokaal geen grote verzameling gegevens opgeslagen hoeven te worden. Bij beide werkwijzen is van belang om goed na te denken over wat beoogd wordt met de Big Data analysis, welk probleem wordt aangepakt. Binnen Webvoyager wordt vrij gericht gewerkt, door bijvoorbeeld specifiek te zoeken naar kinderporno of foto's van vuurwapens.

De werkgroep die de toepassing van webcrawlers binnen de politie evalueerde maakte in het verlengde hiervan de terechte opmerking dat internetsurveillance niet het zelfde is als klassieke surveillance:[53]

---

[51] http://www.cloudera.com/content/cloudera/en/products-and-services/cdh/impala.html

[52] Prive, e-mail communicatie, "on file with author", zoals Amerikanen plachten te zeggen.

[53] Uit niet openbaar rapport, ibidem.

> "het rondkijken op internet verschilt dusdanig van surveilleren dat de term internetsurveilleren onjuist is en de verkeerde connotaties oproept. De regels die de fysieke surveillance reguleren zijn ongeschikt voor toepassing op internet."

In vergelijking met de fysieke wereld is de hoeveelheid informatie op internet onvergelijkbaar veel groter en bovendien in veel gevallen persistent. Als in de openbare ruimte wordt rondgekeken kunnen agenten tijdens surveillance op verdachte gedragingen reageren. Die nemen ze op het moment van handelen waar. Op internet is een veelheid aan informatie van en over personen te vinden, veel meer en rijkere informatie dan ter plekke kan worden geconstateerd.

Het valt buiten het kader van dit rapport, maar de toekomst is in dit opzicht boeiend, namelijk op het moment dat de opsporing met Google glass of vergelijkbare devices worden uitgerust, kan de fysieke opsporing naadloos overlopen in internetopsporing. Om de verschillen tussen klassieke surveillance en internetsurveillance te verduidelijken, zullen we kort ingaan op de aard van informatie op internet en de verschillen met de fysieke omgeving.

6.3.1.1 Achtergrond informatie verzamelen op internet

Wat betreft informatie op internet die op een persoon betrekking heeft geldt dat deze lang niet altijd van deze persoon afkomstig is en zelfs als dit wel zo is, is niet op voorhand duidelijk of de betreffende persoon wist of wilde dat deze informatie door de overheid in het algemeen en de opsporing in het bijzonder wordt gebruikt. Als samenleving moeten we onze weg nog vinden in wat wij toelaatbaar vinden om te gebruiken en wat niet. De overheid moet hierbij een voortrekkersrol vervullen.

6.3.1.2 Openbare informatie: mogelijk, dus toelaatbaar?

De technologie bepaalt de mogelijkheden, juristen houden zich bezig met de normatieve vraag welke technische mogelijkheden binnen een democratische rechtstaat juridisch toelaatbaar zijn. Het idee leeft

dat informatie die op internet beschikbaar is voor het algemene publiek door een ieder gebruikt mag worden. Dit is in beginsel ook zo. Er kunnen natuurlijk voorwaarden worden gesteld aan het verdere gebruik van deze informatie, bijvoorbeeld op grond van het auteursrecht. Maar ook bij vrijelijk beschikbare informatie is het lastig om aan te geven welk gebruik daarvan geoorloofd is.

Er wordt onderscheid gemaakt tussen open bronnen op internet en alleen met wachtwoord toegankelijke of anderszins beperkt beschikbare, niet openbare informatie. Openbare of vrij beschikbare informatie is uiteraard geen garantie voor de juistheid ervan en het betekent ook niet dat degene over wie deze informatie gaat hier weet van heeft. Technisch is het zondermeer mogelijk deze openbare informatie te gebruiken, maar vanuit juridisch en privacy perspectief moet hier terughoudend en tenminste kritisch mee worden omgegaan. Ook bij publieke informatie zijn de regels van verwerking van persoonsgegevens van toepassing, aldus WP29:[54]

> "it is important to note that any information relating to an identified or identifiable natural person, be it publicly available or not, constitutes personal data. Moreover, the mere fact that such data has been made publicly available does not lead to an exemption from data protection law."

Dit geldt in versterkte mate als informatie die op internet beschikbaar is gekoppeld wordt aan binnen een overheidsorganisatie aanwezige informatie. Bij het op deze manier verrijken van overheidsinformatie over burgers is sprake van verwerking van persoonsgegevens en moet dus een duidelijk doel bepaald worden waarom de gegevens verwerkt worden. Dit doel moet proportioneel zijn. Dit staat op gespannen voet met de mogelijkheden die Big Data analysis biedt. We gaan niet zover dat wij stellen dat het onder alle omstandigheden ongeoorloofd is om op deze wijze exploratief

[54] WP29, Opinion 03/2013 on purpose limitation, 2 april 2013

informatie te analyseren, maar op het moment dat je bijvoorbeeld al het telefoonverkeer monitort is dit per definitie niet proportioneel.

6.3.1.3 Profielen en "hinderlijk volgen"

Door het intensief zoeken van allerhande informatie over een persoon kan een profiel verkregen worden dat is opgebouwd uit elementen die op zichzelf vrijelijk beschikbaar zijn. Juridisch is tegen deze werkwijze niet gemakkelijk iets in te brengen, maar de combinatie van al deze elementen vormt niet, zeker niet in het licht van onbewust of ongewenst op internet aanwezige informatie, per se een betrouwbare weergave van een persoon. Zeker als deze activiteiten door de overheid plaatsvinden kan de volgende kanttekening worden geplaatst naar analogie van het volgen van personen in de fysieke wereld.

Als iemand de gordijnen niet heeft gesloten, is het toegestaan naar binnen te kijken, juridisch gezien. Hier zijn echter grenzen. Als je langsloopt kun je een blik naar binnen werpen, maar open gordijnen betekent niet dat het daarom geoorloofd is om buitenaf met grote regelmaat of gedurende langere tijd alles te volgen wat zich binnenshuis afspeelt. Hetzelfde geldt voor gedrag in de openbare ruimte, waarvoor het strafrecht de bepaling hinderlijk volgen (art. 426bis Sr) kent. Als iemands reilen en zeilen in de fysieke, openbare ruimte nauwlettend in de gaten gehouden wordt, dan kan dit dus zelfs tot strafrechtelijke sanctionering leiden. Een vorm van hinderlijk volgen op internet is recentelijk gesanctioneerd in de vorm van de veel besproken cookie-regelgeving, die "track en tracing"-cookies niet toestaat tenzij de gebruiker met de plaatsing van dergelijke cookies heeft ingestemd.

De strafbaarstelling van hinderlijk volgen en stalking heeft niet alleen betrekking op de fysieke aanwezigheid maar ook op de informatiegaring. Strafrechtelijk is stalking gesanctioneerd in art. 285b Sr.

> 1. Hij, die wederrechtelijk stelselmatig opzettelijk inbreuk maakt op eens anders persoonlijke levenssfeer met het oogmerk die ander te dwingen iets te doen, niet te doen of te dulden dan wel vrees aan te jagen wordt, als schuldig aan belaging, gestraft met een gevangenisstraf van ten hoogste drie jaren of een geldboete van de vierde categorie.

De inbreuk op de persoonlijke levenssfeer, wat het nauwlettend iemands sporen volgen op het internet in zekere zin is, is strafbaar indien iemand wordt gedwongen iets te doen, te laten of toe te staan dan wel *"vrees wordt aangejaagd".* Dit komt het dichtst bij uitgebreid informatie over iemand verzamelen. Het kan beangstigend zijn te constateren hoeveel iemand over je weet. Toch zal dit enkele gevoel alleen in extreme gevallen tot strafrechtelijke vervolging kunnen leiden.

Ook de civielrechtelijke stalking (onrechtmatige daad) heeft twee kanten, namelijk het hinderen van een persoon en het inmengen in de persoonlijke levenssfeer van deze persoon. Op internet kan iemand zichtbaar een ander lastig vallen, maar het is net als bij het opbouwen van een profiel ook mogelijk dit te doen zonder sporen achter te laten. Op het internet zal het volgen van iemands reilen en zeilen minder snel opvallen dan in de fysieke wereld. Dat het om die reden ook is toegestaan, zeker als overheid, is niet zondermeer het geval. Ook hier geldt dat het van belang is dat de overheid kritisch nagaat *waarom* er van dergelijke informatie gebruik gemaakt wordt en met welk doel de persoonsgegevens verwerkt worden.

6.3.1.4 Informatie via sociale media

Op sociale media is veel informatie over personen te vinden. Het browsen door sociale media kan allerhande informatie opleveren over een persoon, wat deze meemaakt en met wie deze contacten heeft. Gebruikers zijn zich niet altijd bewust van het bereik van de informatie die op sociale media als Facebook en Twitter wordt geplaatst. Ook op deze "bladerende" wijze kan handmatig een vrij

uitgebreid profiel van iemand worden gecreëerd, laat staan als deze analyse geautomatiseerd plaatvindt. Soms is op sociale media aanwezige informatie afgeschermd binnen het betreffende platform, maar wordt deze wel ontsloten via zoekmachines of is beschikbaar via andere diensten.

Zo kan bij een Facebook profiel worden aangegeven dat het profiel niet gevonden mag worden wanneer op iemands naam gezocht wordt binnen Facebook. In dat geval komt een betreffend profiel binnen Facebook dan ook niet terug in de zoekresultaten. In een door ons geconstateerd geval[55] is echter als via Google op de naam gezocht wordt het Facebook profiel zelfs de tweede treffer. Dit is een illustratie van informatie die op internet beschikbaar is waarvan de betrokken persoon zelf expliciet heeft aangegeven dat deze niet beschikbaar zou moeten zijn.

Een ander voorbeeld eveneens gerelateerd aan Facebook is het tonen van foto's die op prive staan, dus alleen bedoeld zijn voor vrienden. Tinder App is een op moment van schrijven populaire 'flirt'-smartphone toepassing die het mogelijk maakt om via location-based services te achterhalen of er iemand in de nabijheid is die binnen het voorkeursprofiel van de gebruiker valt. Uit technisch praktijkonderzoek in november 2013 door studenten van de master Internet, Intellectuele eigendom en ICT[56] bleek dat via deze App foto's met andere gebruikers werden gedeeld waarvan was aangegeven in Facebook dat deze niet gedeeld mochten worden buiten de kring van Facebook vrienden.

Er is dus op verschillende manieren, veelal indirect, aan informatie van sociale media te komen waarvan een gebruiker expliciet heeft aangegeven deze niet te willen delen. Dergelijke informatie kan een van de grondstoffen zijn waar Big Data analyse op draait.

[55] Het was niet een test/experiment, maar we kwamen hier bij toeval achter.

[56] Sandy Pronk, Samuel Wiegerinck en Gregory Van Zetten, zie http://lodder.cli.vu/smartproject

Over het internet werd in de Memorie van Toelichting bij Computercriminaliteit II in algemene zin gesteld:[57]

> Vervolgens kunnen politie-ambtenaren als ieder ander rondkijken in de digitale wereld en kennis nemen van de voor een ieder raadpleegbare informatie. Daarvoor is niet vereist dat zij een verdenking van een strafbaar feit hebben. Evenmin behoeven zij hun hoedanigheid van opsporingsambtenaar bekend te maken. Zoals de politie, al dan niet in burger, op straat mag surveilleren en rondkijken, zo mag een rechercheur vanachter zijn computer hetzelfde doen op Internet.

Deze stelling is met name door de opkomst van sociale media niet langer houdbaar. Een politie-ambtenaar mag bijvoorbeeld niet zonder verdenking in de uitoefening van zijn functie uitgebreid allerlei profielen van sociale netwerken bekijken.

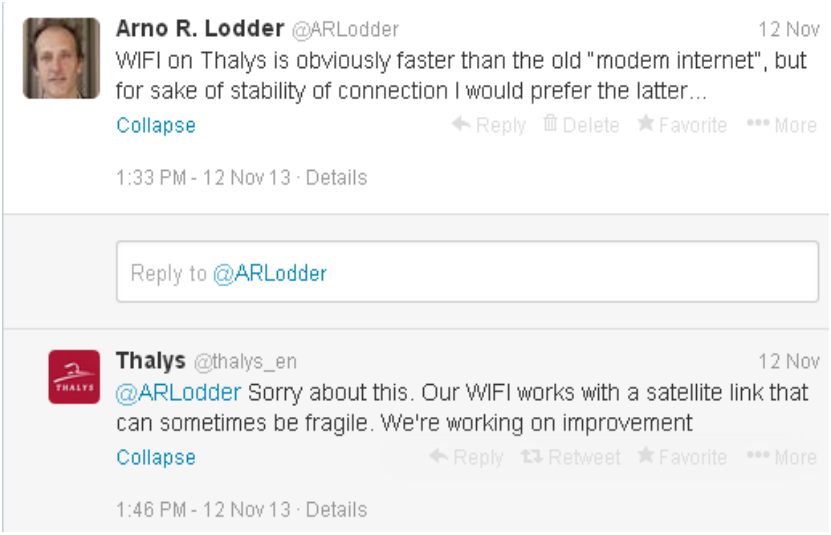
6.3.2 Private software

Opsporingsinstanties kunnen zelf gegevens verzamelen, maar kunnen eveneens gebruik maken van marktpartijen. In Nederland kan bijvoorbeeld gebruik gemaakt worden van bedrijven zoals Coosto, een bedrijf dat zichzelf profileert als een 'marktleider' op het gebied van social media monitoring en webcare. Volgens de site van Coosto kunnen cliënten 'onbeperkt zoeken naar een onbeperkte hoeveelheid resultaten.' Onbeperktheid overheerst, zo blijkt uit de manier waarop het bedrijf zichzelf profileert. Coosto verzamelt volkomen ongericht zoveel mogelijk informatie van sociale media en sites waar input van gebruikers voor bedrijven relevant kan zijn zoals Radar, Kelkoo en consumentensites. Al deze informatie wordt lokaal opgeslagen en kan gebruikt worden om aan klanten bijna real time aan te geven wie, wat, op welk moment over een bedrijf of instelling

[57] *Kamerstukken* 1998/99, 26 671, nr. 3, p. 35.

heeft gezegd. De service die Coosto biedt, is dus het signaleren van relevante communicatie binnen sociale media door offline opgeslagen content te doorzoeken.

Coosto is een leverancier voor in totaal 9 ministeries, waaronder het ministerie van Veiligheid en Justitie. Het is dus een programma dat wordt gebruikt om het image van een bedrijf in de sociale media te monitoren. Positieve berichten kan op ingesprongen worden door deze bijvoorbeeld te delen met volgers en negatieve geluiden kan op gereageerd worden. Als bijvoorbeeld via Twitter de bedrijfsnaam wordt genoemd, dan wordt daar een alert verzonden. Het kan goed een ander programma dan Coosto zijn geweest, maar onderstaande reactie kan alleen zo snel (binnen 15 minuten) plaatsvinden zonder dat een bedrijf op een Twitter manier genoemd wordt, hier dus @thalys_en, omdat continue Tweets gescand worden op voor het betreffende bedrijf voorkomende relevante termen.



In antwoord op Kamervragen stelt minister Blok dat politie en ministeries uitsluitend gebruik van de zoektechnologie van Coosto maken en zelf het onderzoek uitvoeren:[58]

> De politie en de ministeries maken alleen gebruik van de zoektechnologie van Coosto om binnen sociale netwerken te zoeken in openbare bronnen. Voor de ministeries is dit een activiteit vergelijkbaar met het maken van een knipselkrant. De ministeries voeren echter zelf het onderzoek uit.

De parallel met een knipselkrant is ongelukkig. De mogelijkheden van dergelijke software strekken veel verder en stellen de gebruikers in staat om bij wijze van spreken hele boeken over een persoon of onderwerp samen te stellen. Er is een spanning tussen enerzijds vanuit Big Data perspectief gewenste onbeperkt verzamelen door opsporingsinstanties en anderzijds te betrachten terughoudendheid vanuit legitimiteit en geloofwaardigheid alsmede het respecteren van de persoonlijke levenssfeer.

6.3.3 Ongericht versus gericht: drie scenario's

Over het algemeen zijn er voor opsporingsinstanties drie scenario's te bedenken met betrekking tot webcrawlen. Allereerst het random verzamelen van data geheel zonder beperkingen en zonder duidelijk doel. Dit kan voor de opsporing interessant zijn omdat ongericht verzamelen niet eerder ontdekte patronen of resultaten kan opleveren. Voor opsporingsdoeleinden is het volledig ongericht binnenhalen van informatie niet te rijmen met werken op basis van verdenkingen en staat daarnaast op gespannen voet met de proportionaliteitstoets die bij privacy-inbreuken altijd moet worden meegenomen. De verdenking hoeft niet gerelateerd te zijn aan een verdachte, maar kan ook bestaan uit objecten (zoeken naar

[58] Minister Blok 5 november 2013. Antwoorden op de vragen van het Kamerlid Van Toorenburg (CDA), die zijn gesteld bij brief d.d. 20 september jl., met kenmerk 2013Z17854.

vuurwapenhandel, kinderporno) of bepaalde dadereigenschappen (dreigend taalgebruik).

Andrejevic (2014) geeft aan hoe in het Verenigd Koninkrijk in 2012 door de politie werd aangekondigd dat ze over gingen van een systeem van targeted surveillance naar generalized surveillance. Hoewel dit tot privacy zorgen leidde geeft Andrejevic aan hoe de Britse politie benadrukte dat hun aandacht vooral uitging naar patronen en niet naar de inhoud van, bijvoorbeeld, emailberichten. De benadering is dus *inductive*, verzamel zoveel mogelijk informatie met als doel om te kijken of patronen te voor schijn komen (Andrejevic 2014). Het verzamelen van zoveel mogelijk data over zoveel mogelijk mensen wordt daardoor bijna een doel op zich. Zoals Andrejevic aangeeft betekent dit niet dat iedereen als verdachte gekenmerkt wordt door de politie, maar om verdachten te vinden moet iedereen object kunnen zijn van surveillance. Vanuit een privacy perspectief is het ontbreken van beperkingen ongewenst.[59]

Het tweede scenario betreft daarom het introduceren van enkele beperkingen om de verzameling een richting te geven. Dit is ons inziens altijd aan te raden, voorafgaand aan het verzamelen nadenken over waarnaar gezocht wordt. Deze insteek wordt bij Webvoyager ook nadrukkelijk gevolgd.

De derde optie wordt nog specifieker door beperkingen te introduceren op het gebruik van de verzamelde gegevens. Deze drie scenario's zijn te plaatsen op een spectrum, van algemeen naar specifiek, maar ook op het gebied van proportionaliteit. Data

[59] Zie ook het Europese FP7-project VIRTUOSO over open source intelligence, waarin de juridische analyse erg gericht is op prototype, auteursrecht, aansprakelijkheid en slechts zijdelings op privacy. De laatste van 10 aanbevelingen is meest relevant en luidt: "10. Apply privacy/security by design by restricting the functionalities of the end product(s) as much as possible (i.e., make them minimally invasive for human rights and intellectual property rights) while ensuring they can serve their intended purpose of open-source intelligence by public authorities for public security."

verzamelen zonder enige beperking kan gezien worden als disproportioneel. Hoewel wellicht de diepgang van de privacy inbreuk ontbreekt, is de breedte van de inbreuk wel een factor om rekening mee te houden. In die zin kan Big Data in de opsporing privacy als gemeenschappelijk goed aantasten, meer nog dan privacy als individueel recht.

7 Uitgangspunten Big Data, in het bijzonder vanuit het oogpunt van verwerking persoonsgegevens

'Big Data' is wat betreft de auteur toch vooral een variant op 'big brother', met onvoorzienbare consequenties van grootschalige koppeling van gegevens en de daaruit wellicht voortvloeiende kennis over gedrag van mensen – kennis die eenvoudig zelfkennis kan overstijgen. Oftewel: straks weten anderen veel meer over ons dan wijzelf, op basis van ons dagelijkse gedrag.

Mommers (2014)

Voor bedrijven zowel als overheden is het van belang te realiseren dat privacy niet een doorsnee recht is. Privacy is een grondrecht, een fundamenteel recht, een mensenrecht. Neelie Kroes verwoordde het in februari 2014 duidelijk:[60]

> "People – including me - sometimes about talk about our "digital rights". But I don't think that's quite right. These are not digital rights, nor online rights: they are fundamental rights, and they apply just as much online as off. Whether it is privacy, or freedom of speech, or consumer protection. New technology can enhance our humanity: it should not override our human rights. The massive scale of online spying shows how technology can be used for ill. Invading privacy, invading fundamental rights, eroding trust in the online world: and in our governments. This is totally unacceptable."

[60] N. Kroes, A secure online network for Europe, *Cyber security conference*, Brussels, 28 February 2014

Iets daarvoor, in juni 2013, liet Obama rond de NSA affaire een vergelijkbaar geluid horen:[61]

> "When I came into this office, I made two commitments that are more than any commitment I make: number one, to keep the American people safe; and number two, to uphold the Constitution. And that includes what I consider to be a constitutional right to privacy and an observance of civil liberties"

Big Data analysis kan inbreuk maken op privacy. Niet ieder inbreuk is ongeoorloofd, want als er een wettelijke grondslag is dan mag in strijd met privacy gehandeld worden. De wet bescherming persoonsgegevens is een uitwerking van deze uitzondering. Ook hier moet het fundamentele recht in het achterhoofd gehouden worden, omdat zelfs bij legitieme verwerkingen, bijvoorbeeld met toestemming van de betrokkene, steeds de vraag moet worden gesteld of het handelen wel proportioneel is.[62]

De privacy zorgen die over Big Data analysis bestaan zijn door de WP 29 als volgt verwoord:

> - the sheer scale of data collection, tracking and profiling, also taking into account the variety and detail of the data collected and the fact that data are often combined from many different sources;
> - the security of data, with levels of protection shown to be lagging behind the expansion in volume;
> - transparency: unless they are provided with sufficient information, individuals will be subject to decisions that they do not understand and have no control over;

---

[61] http://www.whitehouse.gov/the-press-office/2013/06/07/statement-president

[62] Hoge Raad 9 September 2011, ECLI:NL:HR:2011:BQ8097

- inaccuracy, discrimination, exclusion and economic imbalance (as will be discussed further below); and
- increased possibilities of government surveillance.

Kort gezegd komt het neer op de enorme omvang en gedetailleerdheid van de te verwerken data, de beveiliging, de inzichtelijkheid en mogelijke inaccuraatheid. De zorgen over toegenomen mogelijkheden van surveillance wordt ook door de Verenigde Naties gedeeld. De algemene vergadering van de Verenigde Naties heeft op 19 december 2013 een resolutie vastgesteld inzake *Privacy in the digital age*.[63] In deze resolutie worden zorgen geuit over de mogelijke schending van mensenrechten door surveillance en het verzamelen van persoonsgegevens:

> "Deeply concerned at the negative impact that surveillance and/or interception of communications, including extraterritorial surveillance and/or interception of communications, as well as the collection of personal data, in particular when carried out on a mass scale, may have on the exercise and enjoyment of human rights"

Ook worden de lidstaten verzocht de regelgeving en bestaande praktijk te evalueren:

> "To review their procedures, practices and legislation regarding the surveillance of communications, their interception and collection of personal data, including mass surveillance, interception and collection, with a view to upholding the right to privacy by ensuring the full and effective implementation of all their obligations under international human rights law"

[63] http://www.un.org/ga/search/view_doc.asp?symbol=A/C.3/68/L.45/Rev.1

In dit slothoofdstuk geven we op grond van hetgeen naar voren is gekomen in de eerdere hoofdstukken de privacy uitgangspunten bij Big Data Analyse aan, maar eerst wordt kort ingegaan op een andere insteek, namelijk die Lokke Moerel in haar oratie naar voren bracht op 14 februari 2014.

7.1 Big Data Protection van Moerel

De huidige privacy regelgeving legt in de ogen van Moerel te veel beperkingen op aan de Big Data industrie. Zij stelt daarom voor om andere de volgende beginselen uit de in de verordening voorgestelde regelgeving te schrappen:

• purpose limitation principle;
• data minimisation principle;
• storage minimisation principle;
(…)
• right to object to profiling;

Het is wat ambitieus om in de beperkte ruimte die een oratie biedt voor te stellen de meeste, al tientallen jaren bestaande, kernbeginselen van gegevensverwerking te elimineren. Hoewel de spanning van de beginselen met Big Data analyse duidelijk is, hechten wij er vooralsnog wel aan deze beginselen te handhaven. Zij vormen de randvoorwaarden waarmee de inbreuken op het grondwettelijk privacyrecht worden gerechtvaardigd.

Moerel stelt voor om in plaats van bovenstaande en nog enkele andere beginselen het volgende als uitgangspunt te nemen:

> "Extending the "legitimate interest ground" to the processing of all categories of data and further to all phases of the life-cycle of data"

Dit is een nogal beperkte vervanging. De rechtmatige grondslag die nodig is om gegevens te verwerken wil zij uitbreiden tot alle gegevens, niet alleen persoonsgegevens. Zij voegt dus geen enkele beperking toe aan de verwerking boven op de al bestaande rechtmatige grondslag. Deze grondslag is beperkter dan de bestaande, zelfs de toestemming en de wettelijke plicht worden niet als grondslag vermeld:

> "de gegevensverwerking noodzakelijk is voor de behartiging van het gerechtvaardigde belang van de verantwoordelijke of van een derde aan wie de gegevens worden verstrekt, tenzij het belang of de fundamentele rechten en vrijheden van de betrokkene, in het bijzonder het recht op bescherming van de persoonlijke levenssfeer, prevaleert."

De vraag is waar tegen de noodzakelijkheid wordt afgezet als het doel voor de verwerking niet bepaald hoeft te worden. Ik vermoed dat een bedrijf dat Big Data analyseert al snel zal kunnen betogen waarom dat noodzakelijk voor haar bedrijfsvoering is. Bij overheden zou op dit punt meer terughoudendheid kunnen worden verwacht, hoewel met name de inlichtingen- en veiligheidsdiensten niet erg hebben bijgedragen aan vertrouwen in de overheid op dit punt.

In haar toelichting noemt ze in wezen alleen de bovenstaande grondslag en geeft aan dat de afweging moet plaatsvinden op grond van onder andere een kosten-baten analyse en vanuit een "harm-based" perspectief. Met dit laatste wordt bedoeld dat niet de (grond)rechten als zodanig centraal staan, maar dat gekeken moet worden aan wie en op welke wijze een handeling mogelijk schade toebrengt, of in de woorden van Moerel (p. 26):

> "Any government regulation in the area of data protection needs to balance the interests of organisations (companies and governments) that use personal data against the potential harm such use could cause individuals."

Ze voegt nog wel enkele beperkingen en nadere waarborgen toe waar de afweging aan moet voldoen, maar vooralsnog zijn wij niet direct overtuigd van deze vernieuwing en blijven uitgaan van de reeds behandelde beginselen. Een oratie is echter geen eindpunt maar een begin, dus Moerel heeft de komende de jaren de tijd om duidelijk te maken dat haar standpunt houdbaar is. Inspiratie kan daarbij worden geput uit een in maart 2014 door de van de European Data Protection Supervisor gepresenteerd stuk *Privacy and competitiveness in the age of big data* waar wordt ingegaan op de verhouding tussen privacy, consumentenbescherming en mededinging. Aan het eind worden onderzoekslijnen en andere punten van aandacht geformuleerd:[64]

"A comprehensive response to these challenges requires more time for investigation, reflection and discussion, but might include any or all of the following:
raised awareness among consumers, service providers and regulators of current and future technological developments in relevant markets in the digital economy and the implications for competitiveness, consumer welfare and choice and innovation around privacy-enhancing services;
effective guidance on the application of privacy, competition and consumer protection rules for online services, in particular those promoted as 'free' services, which takes into account the views of customers and competitors and evidence of customer preferences and concerns;
cooperation between authorities in investigation and enforcement, for example in identifying scenarios and possible standards for measuring market power in the digital economy, and consultation on investigations into individual cases; and

[64] Preliminary Opinion of the European Data Protection Supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, March 2014, p. 37-38.

a review of competition legislation for 21st century digital markets, including its interfaces with other areas of law and possibilities for productive interaction with other relevant authorities."

Het laatste woord is hier nog lang niet over gezegd.

7.2 Bepaal te analyseren probleem en specificeer doel voor verwerking

Big data analysis begint met het definiëren van welk probleem onderzocht gaat worden. Dit is ook het moment dat over het doel van de verwerking moet worden nagedacht. Omdat er bij Big Data als vanzelf veel gegevens worden verwerkt zal het, een enkele uitzondering mogelijk daargelaten, ook vrijwel altijd tenminste mede tot de persoon herleidbare informatie betreffen. Zelfs als dat niet zo is, is het verstandig om vanaf het begin na te denken over het doel van de verwerking, omdat er altijd op enig moment binnen het Big Data analyse proces persoonsgegevens verwerkt zullen worden of naar aanleiding van de analyse handelingen verricht worden die een specifieke persoon betreffen.

7.3 Selecteer data en beperk verzamelen

Nadat het probleem gedefinieerd is en het doel gespecificeerd, volgt bij Big Data analysis de selectie van de gegevens. Hier is van belang om niet alles wat mogelijk is te analyseren, maar deze analyse te beperken. Dit brengt niet alleen het data limiteringsbeginsel mee. Ook vanuit het oogpunt van fundamentele rechten is dit een belangrijk punt. Hoewel bij Big Data analysis veelal betere resultaten behaald worden als meer data gebruikt worden, hoeft dit uiteraard niet te betekenen dat in plaats van meer, *alles* wordt meegenomen in de analyse. Dan ontstaat er een probleem met de mensenrechtelijk verankerde proportionaliteit. Een inbreuk op privacy is op grond van het EVRM en het Handvest van de EU immers enkel toelaatbaar als

deze proportioneel is. Dit vraagt per geval om een afweging. Er zijn situaties waarbij het op voorhand duidelijk is dat de maatregel niet proportioneel is. Indien bijvoorbeeld *al* het telefoonverkeer binnen een land, of binnen een stad, wordt gemonitord kan dit nimmer als een proportionele inbreuk worden gekwalificeerd.

7.4 Bewaar niet langer dan noodzakelijk

Als de gegevens geanalyseerd worden vanuit lokaal opgeslagen databases, zoals bijvoorbeeld Coosto doet bij gegevens van social media, moet de informatie niet onbeperkt opgeslagen worden. Hier speelt het gedefinieerde doel van de verwerking een rol. De gegevens moeten zolang bewaard worden als voor het realiseren van het doel noodzakelijk is. Zeker binnen de overheid moet voorkomen worden dat er grote gegevensverzamelingen bewaard worden omdat die mogelijk ooit van pas zouden kunnen komen, maar men nog geen idee heeft waarvoor of wanneer. Terughoudendheid is ook hier op zijn plaats.

7.5 Wees transparant

Een jaar of 10 geleden waren er berichten over een terreurscore die de Amerikanen bijhielden. Deze score deelde mensen in op grond van bepaalde eigenschappen waar terroristische dreiging vanuit ging. Iemand die scheikunde studeert, dat wordt deze studenten ook wel meegedeeld, krijgen wat extra punten vanwege de toegang die zij hebben tot chemische middelen. Er zijn tenminste twee problemen met deze terreurscore. De eerste hangt samen met het vorige uitgangspunt, namelijk dat deze scores 40 jaar werden bewaard. Het andere probleem is het gebrek aan inzichtelijkheid. Dit betekent dat er beslissingen genomen kunnen worden zonder dat een individu begrijpt waarom. Het is de vraag in hoeverre het gebrek aan transparantie bij veiligheidsdiensten terecht is, maar bij de overheid in het algemeen is het van belang duidelijk te zijn over waarom er

precies Big Data analyse wordt toegepast en wat er met de resultaten gebeurt. Dit biedt ook de mogelijkheid om te controleren hoe zij opgenomen zijn in bepaalde systemen, bijvoorbeeld welke profielen de overheid van toepassing acht.

7.6 Beveilig informatie

Privacy en security worden wel als tegenstellingen gepresenteerd, maar in ieder geval in termen van informatiebeveiliging heeft privacy zonder security niet veel nut. Gegevens die voor Big Data analysis worden gebruikt, zeker als dit grote verzamelingen lokaal opgeslagen informatie betreft, moeten zowel organisatorisch als technisch beveiligd zijn. Niet iedereen moet erbij kunnen, en als er toegang wordt verleend moet deze gelogd worden. De technische beveiliging kan nooit perfect zijn, maar zal moeten voldoen aan de stand der techniek. De beveiliging moet bij iedere gegevensverwerking, zeker bij Big Data omdat dit uit de aard grote hoeveelheden gegevens betreft, goed vormgegeven zijn.

7.7 Evalueer de uitkomsten kritisch

De correlaties en patronen die gevonden worden door Big Data analysis moeten altijd kritisch bekeken worden. Het is bekend dat bij gebruik van technologie mensen na verloop van tijd hun kritisch vermogen verliezen en de uitkomsten overnemen zonder daarover te reflecteren. Dit is in ieder geval voor juristen empirisch vastgesteld (Dijkstra 1995), maar geldt vast ook voor andere beroepsgroepen. Hier is dus waakzaamheid geboden. Ook na enkele succesvolle analyses zal steeds op de uitkomsten kritisch gereflecteerd moeten worden.

7.8 Slotopmerking

Een mogelijke richting die is voorgesteld door Mayer-Schonberger& Cukier (2013, p. 174) is om de nadruk te leggen op “accountability” van de gebruiker van Big Data analysis en minder bij de betrokkene:

> Shifting the burden of responsibility from the public to the users of data makes sense

Het is in zekere zin ook de lijn die Moerel voorstelt. Wij zijn er niet direct van overtuigd dat het bedrijfsleven deze verantwoordelijkheid aan kan. In dit hoofdstuk hebben wij enkele uitgangspunten met name ontleend aan beginselen van verwerking van persoonsgegevens toegelicht. Voor de overheid is het zeker van belang om zorgvuldig met Big Data om te gaan en hier goed over na te denken.

8 Conclusie

wij weten alles over hem, maar hij weet niets van zichzelf
Dommering (2008)

Big data analysis heeft twee gezichten. Zo kan men enerzijds zeggen dat kennis macht is, kennissen machtiger en Big Data analysis machtigst, maar anderzijds ook lies, damn lies, Big Data analysis. In dit rapport hebben wij enkele mogelijk Big Data toepassingen binnen justitie verkend en daarbij vooral stilgestaan bij de privacy aspecten daarvan.

Het is onmiskenbaar dat Big Data analysis mogelijkheden biedt. In hoofdstuk 2 is in algemene zin op Big Data en Big Data Analysis ingegaan, en de relevante privacy normen zijn in hoofdstuk 3 behandeld. Wat privacyregulering betreft zitten we in een overgangsfase tussen het regime zoals dat in hoofdzaak voor het ontstaan van het internet is ontwikkeld (Wbp) en de nog niet vastgestelde EU Privacy verordening en voor de opsporing de corresponderende richtlijn. Duidelijk is dat uitgangpunten als doelbinding en dataminimalisatie zich moeizaam laten verenigen met Big Data analysis. Hoewel er in toekomstige privacy regelgeving onder andere aandacht is voor profilering, bieden ook de voorgestelde regels niet altijd de gewenste waarborgen voor de veelal indirecte verwerkingen en inbreuken die bij Big Data analysis kunnen plaatsvinden. Op verschillende plaatsen in het rapport benadrukken we dan ook dat de overheid zowel voorafgaand aan Big Data analysis kritisch moet nadenken over wat nu precies beoogd wordt, als na afloop kritisch moet reflecteren op de uitkomsten.

In hoofdstuk 4 zijn enkele mogelijke en bestaande toepassingen binnen de rechtspraak besproken. Door de toenemende en op niet al te lange termijn volledige digitalisering leent de rechtspraak zich voor Big Data toepassingen. Wederom moet hier goed nagedacht worden over wat men hiermee beoogt. Zo zal Big Data analysis van vonnissen

zonder duidelijk doel voor ogen tot uitkomsten kunnen leiden waar voorzover juist men niet noodzakelijkerwijs over geïnformeerd wil zijn. Ver doorgevoerde managementinformatie kan op gespannen voet met de onafhankelijkheid van de rechter komen te staan. Het is niet te voorkomen dat bedrijven allerhande analyses gaan toepassen zoals nu in de Verenigde Staten van Amerika gebeurt om advocaten in patent-zaken te adviseren. Dergelijke ontwikkelingen moeten zorgvuldig geëvalueerd worden, met name omdat kenmerkend voor Big Data analysis is dat uit de data bepaalde conclusies volgen (correlatie) maar dat de vraag naar waarom dit zo is (onder andere causatie) niet altijd duidelijk is en in sommige gevallen ook niet wordt. De vraag is wat de betekenis van dergelijke conclusies in het juridische veld is. Het voorspellen van uitkomsten in rechtszaken is daar een voorbeeld van. Goed denkbaar dat dit vrij accuraat gebeurt, maar dat de factoren die daarvoor bepalend zijn juridisch gezien onwenselijk zijn. Stel dat de verdachte significant minder straf krijgt als hij blond haar heeft, heb je dan als advocaat de plicht je cliënt te adviseren zijn haar te blonderen? Een punt waar de rechtspraak inclusief de betrokken advocaten en partijen zich tenslotte bewust van moeten zijn is dat zij onderwerp van Big Data analysis door derden kan worden, zowel bedrijven als overheden. Volledige openheid van procesinformatie kan zo ook duidelijk nadelen meebrengen.

In hoofdstuk 5 zijn mogelijke en bestaande toepassingen binnen de opsporing besproken. Binnen de opsporing is het gebruik van informatie over gebeurtenissen en personen noodzakelijk om uiteindelijk tot een veroordeling te komen. Wij zijn ingegaan op zogenaamde webcrawlers alsmede predictive policing. Op het moment dat er gerichte Big Data analysis wordt toegepast, vanwege een verdachte persoon of verdachte objecten, kunnen deze activiteiten in beginsel onder de bestaande bevoegdheden worden uitgevoerd. Lastiger wordt het als op grote schaal informatie wordt opgeslagen om eventueel op een later moment te gebruiken. Of wanneer ongericht naar informatie wordt gezocht. Zeker dit laatste

kan niet zonder nieuwe bevoegdheden, maar we zijn er geen voorstander van om deze te creëren. Hoewel er een tendens is om grote hoeveelheden ongericht verzamelde informatie te bewaren (verkeersgegevens, camerabeelden) is vooralsnog niet overtuigend aangetoond dat hiermee resultaten worden bereikt. Het ongebreideld, ongericht verzamelen van informatie laat zich lastig verenigen met het op verdenkingen gebaseerde opsporingswerk. Wij raden dergelijk gebruik binnen de opsporing af.

Voortbouwend op hoofdstuk 2-5, volgde in hoofdstuk 6 het antwoord op de probleemstelling.

> Welke juridische en met name privacyrechtelijke uitgangspunten dienen in acht te worden genomen bij de inzet Big Data toepassingen binnen het domein Veiligheid en Justitie teneinde de mogelijkheden die deze technologie biedt optimaal te benutten?

De beschreven uitgangspunten zijn achtereenvolgens:

1. Bepaal te analyseren probleem en definieer doel voor verwerking
2. Selecteer data en beperk verzamelen
3. Bewaar niet langer dan noodzakelijk
4. Wees transparant
5. Beveilig informatie
6. Evalueer de uitkomsten kritisch

9 Literatuur

Andrejevic, M. (2014). “Surveillance in the Big Data Era” in Emerging Pervasive Information and Communication Technologies (PICT), pp. 55 - 69.

Bachner, J. (2013). *Predicting crime with Big Data.* http://www.governmentalstudies.com/govstud/2013/3/7/predicting-crime-with-big-data.html

Big Data and Analytics: Seeking Foundations for Effective Privacy Guidance. A Discussion Document, February 2013, Centre for Information Policy Leadership

Bloem, J. e.a. (2013), *Privacy, technologie en de wet. Big Data voor iedereen door goed design*, Sogeti.

Boonk, M. & A.R. Lodder (2006), Regulating Website Access for Automated Means Such as Search Bots and Agents: Property or Contract? *Contemporary Issues in Law*, Vol. 7, No. 4, pp. 360-374, 2005/2006

boyd, D. and Crawford, K. (2011), Six Provocations for Big Data (September 21, 2011). *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, September 2011

Crawford, K. and Schultz, J. (2014), Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms). *Boston College Law Review*, Vol. 55, No. 1. http://ssrn.com/abstract=2325784

Diebold, F.X (2012*), On the Origin(s) and Development of the Term 'Big Data' (September 21, 2012).* PIER Working Paper No. 12-037.

Dijk, F. van & R. van den Hoogen (2014), Digitale toegankelijkheid van Rechtspraak. Trends en ervaringen, in: S. van der Hof, A.R. Lodder & G.J. Zwenne, *Recht en Computer*, Deventer: Kluwer, p. 283-298.

Dijkstra, J.J. (1995), Dijkstra, J.J. (1995). The influence of an expert system on the user's view: How to fool a lawyer. *New Review of Applied Expert Systems*, 1, 123-138.

Dommering, E. (2008), *‘Gevangen in de waarneming’. Hoe de burger de communicatiemiddelen overnam en zelf ook de bewaking ging verzorgen* (afscheidsrede UvA), Otto Cramwinckel Uitgever.

Fayyad, U., Piatetsky-Shapiro G. & Smyth P. (1996), “The KDD process for Extracting Useful knowledge from volumes of data”, *Communications ACM*, 39(11), pp. 27-41.

Ferguson, A. G. (2012). Predictive Policing and Reasonable Suspicion. *Emory LJ*, *62*, 259.

Greengard, S. (2012). Policing the future. *Communications of the ACM*, *55*(3), 19-21.

Hilbert, M. (2013), *Big Data for Development: From Information- to Knowledge Societies* (January 15, 2013). Available at SSRN: http://ssrn.com/abstract=2205145

Hildebrandt, M. (2012), The dawn of a critical transparency right for the profiling era, in: J. Bus *et al.* (eds.) *Digital Enlightenment Yearbook* 2012, p. 41-56.

Knaap, P. van der & R. van den Broek (2000), Recht van spreken. Een resultaatgericht sturingsmodel voor de rechterlijke macht, *Bestuurskunde*, Jrg. 9, nr. 7, p. 313-325.

Koot, M.R. (2012), *Measuring and predicting anonymity* (diss Amsterdam UvA).

Koops, B.J. (2011), Forgetting Footprints, Shunning Shadows. A Critical Analysis Of The "Right To Be Forgotten" In Big Data Practice, *Script-ed,* pp.229-256

Lerman, J. (2013). Big Data and Its Exclusions. *Stanford Law Review Online*, 66, 55.

Lodder, A.R. (2001), Het strafmaatsysteem het BOS, *R&EM* 2001/3, p. 54-57.

Lodder, A.R. & A. Oskamp (2001), The Netherlands: theoretical perspective, in: A.R. Lodder, A. Oskamp & A.H.J. Schmidt (eds.), *IT support of the Judiciary in Europe* (ITeR no. 43), Den Haag: SDU.

Lodder, Arno R. (2013), Ten Commandments of Internet Law Revisited: Basic Principles for Internet Lawyers *Information & Communications Technology Law*, Vol. 22, Issue 3, http://ssrn.com/abstract=2343486

Mayer-Schonberger, V. & K. Cukier (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think,* Eamon Dolan/Houghton Mifflin Harcourt

Mommers, L. (2014), Over regulering tussen Recht en ICT, in: S. van der Hof, A.R. Lodder & G.J. Zwenne (red.), *Recht en Computer*, Deventer: Kluwer, p. 47-64.

Nissan, E. (2013), Legal evidence and advanced computing techniques for combatting crime: an overview, *Information & Communications Technology Law*

Nolan, Richard (1973). "Managing The Computer Resource: A Stage Hypothesis". *Communications of the ACM* 16 (4): 399–405.

Perry, W. L., McInnis, B., Price, C. C., Smith, S. C. & John S. Hollywood (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Rand Corporation.

Piaget, J. (1970). Piaget's theory. In P. H. Mussen, (Ed.), *Carmichael's handbook of child development* (pp. 703-732). New York: Wiley.

Polonetsky, J., & Tene, O. (2013). Privacy and Big Data: Making Ends Meet. *Stanford Law Review Online*, *66*, 25.

Richards, N.M. & J.H. King (2013), Three Paradoxes of Big Data (September 3, 2013). 66 *Stanford Law Review Online* 41.

Richards, N.M. & J.H. King (2014), Big Data Ethics. *Wake Forest Law Review*, 2014. http://ssrn.com/abstract=2384174

Ross, J.W., C.M. Beath & A. Quaadgras (2013), You May Not Need Big Data After All, *Harvard Business Review* December 2013.

Schermer B.W. (2011), The Limits of privacy in automated profiling and data mining, *Computer law & security report* 27(7): 42-52.

Schwartz, P.M. (2010), *Data Protection Law and the Ethical Use of Analytics*, Centre for Information Policy Leadership

Sloan, R.H. & R. Warner (2013)*, Big Data and the 'New' Privacy Tradeoff (August 5, 2013).* Chicago-Kent College of Law Research Paper No. 2013-33.

Stranieri, A. & J. Zeleznikow (2006), Knowledge Discovery from Legal Databases – using neural networks and data mining to build legal decision support systems, in: A.R. Lodder & A. Oskamp (eds.), *Information Technology & Lawyers,* Springer, p. 81-117

Tene, O. & J. Polonetsky (2013), Big Data for All: Privacy and User Control in the Age of Analytics (September 20, 2012). 11 *Northwestern Journal of Technology and Intellectual Property* 239 (2013)

Tufekci, Z. (2013), *Big Data: Pitfalls, Methods and Concepts for an Emergent Field* (March 7, 2013). Available at SSRN: http://ssrn.com/abstract=2229952

Uijttenbroek, E.M., Lodder, A.R., Klein, M.C.A., Harmelen, F.A.H. van, Wildeboer, G. & Sie, R.L.L. (2008). Retrieval of Case Law to provide laymen with information about liability. In P. Casanovas, G. Sartor, N. Casellas & R. Rubino (Eds.), *Computable Models of the Law* (LNAI 4884), Springer, p. 291-310.

Twee werelden; you only lives ones. Hoofdrapport "Project X" Haren (maart 2013) en 3 Deelrapporten "Project X" Haren (maart 2013).

Viergever, L. & J. Koëter (2012), 'Is onze privacyregelgeving "Big Data proof?"', *Tijdschrift voor Internetrecht* 2012/6.

Zittrain, J. (2008), *The Future of the Internet and How to Stop It*. Yale University Press.

Internet en overige bronnen

Big Data for Development: Challenges & Opportunities. UN Global Pulse (mei 2012), http://www.unglobalpulse.org.

Bloem, J. e.a. (2012), 'Helderheid creeren met Big Data'*,* Sogeti VerkenningsInstituut Nieuwe Technologie VINT (augustus 2012), http://vint.sogeti.com/download-big-data-reports/

Cherry, S. (2013), 'Can Big Data Win Your Next Court Case?', *IEEE Spectrum*, 31 mei 2013, http://bit.ly/1a0qpA4.

Doorenbosch, T. (2013), 'Big Data helpt criminelen vangen', *AutomatiseringGids* 19 februari 2013, http://www.automatiseringgids.nl/nieuws/2013/08/big-data-helpt-criminaliteit-opsporen.

Fischer, D. (2013), 'Backpacker stripped of tech gear at Auckland Airport', *The New Zealand Herald* 12 December 2013, http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=11171475.

Friend, Z. (2013), 'Predictive Policing: Using Technology to Reduce Crime, *The Federal Bureau of Investigation* 9 april 2013, http://www.fbi.gov/stats-services/publications/law-enforcement-bulletin/2013/April/predictive-policing-using-technology-to-reduce-crime.

Geradts, F. (2013), 'Big Data: wat kan, wil en mag je er allemaal mee?', *Emerce* - 26 september 2013, http://www.emerce.nl/cases/leidt-big-data-tot-big-dating.

Goldberg, N.M. & Miller, M.W. (2011), The practice of law in the age of 'Big Data', *National Law Journal* 11 april 2011, http://bit.ly/Mkmo4Q.

Grimes, S. (2013), Big Data: Avoid 'Wanna V' Confusion, *InformationWeek* 8/7/2013

Hellerstein (2008), The Commoditization of Massive Data Analysis, *Data* 19 november 2008, http://strata.oreilly.com/2008/11/the-commoditization-of-massive.html

Kalil, T. & Zhao. F. (2013), 'Unleashing the Power of Big Data', *The White House* (18 april 2013), http://www.whitehouse.gov/blog/2013/04/18/unleashing-power-big-data.

Lampitt, A. (2013),' The real story of how Big Data analytics helped Obama win', *Info World* 14 februari 2013, http://www.infoworld.com/d/big-data/the-real-story-of-how-big-data-analytics-helped-obama-win-212862?page=0,1.

Lim, G. (2013) 'Courts and Big Data', *Innovating Justice Forum* 3 september 2013, http://www.innovatingjustice.com/blogs/big-data.

Maygar, C. (2013), *'*How Big Data analysis helped President Obama defeat Romney in 2012 Elections'. *Bosmol Social Media News* 8 februari 2013, 2013, http://bosmol.com/2013/02/how-big-data-analysis-helped-president-obama-defeat-romney-in-2012-elections.html.

Mehta, A. (2011), Big Data: Powering the next industrial revolution, Tableau, raadpleegbaar via http://isites.harvard.edu/

Nicklaus, D. (2013), *Juristat goes from Startup Weekend to startup sensation*, 15 november 2013, http://bit.ly/LJNUYT.

Nieborg, D.B. (2013), 'Hoe Obama zijn herverkiezingen won', *inGovernment* 7 oktober 2013, http://www.ingovernment.nl/artikelingovernment/hoe-obama-zijn-herverkiezingen-won.

Normandeau, K. (2013), 'Beyond Volume, Variety and Velocity is the is the Issue of Big Data Veracity, *insideBIGDATA* 12 september 2013, http://inside-bigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/.

'Obama administration unveils "Big Data" initiative: announces $200 million in new R&D investments', *The White House* 29 maart 2012, http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf.

Rijmenam, van M. (2013), 'How can Big Data improve the practice of law?', *BigData Startups* 23 augustus 2013, http://www.bigdata-startups.com/how-big-data-can-improve-the-practice-of-law/.

Schepers, M. (2012), 'Big Data: The Next best thing?', *inGovernment* 2012, http://www.ingovernment.nl/artikelingovernment/big-data-next-big-thing.

Schoemaker, R., 'Kroes: Privacy mag niet in de weg staan van Big Data', *Webwereld* 8 november 2013, http://webwereld.nl/big-data/80043-kroes-privacy-mag-niet-in-de-weg-staan-van-big-data.

Thiele, M. (2012), Big Data adoption issues – What's the big deal, Gigacom 26 februari 2012.

Verlaan, D., 'Londen wil slimme wifi-prullenbakken verbannen', *Nu.nl* 12 augustus 2013, http://www.nu.nl/tech/3548562/londen-wil-slimme-wifi-prullenbakken-verbannen.html